AD-A099 106 ARMY TROPIC TEST CENTER APO MIAMI 34004 F/G 5/10
MAGNITUDE ESTIMATION: A NEW METHOD FOR MEASURING SUBJECTIVE TES--ETC(U)
JUN 79 E M HAVERLAND
UNCLASSIFIED USATTC-790601 NL

1 OF 1
AD A
099106

END
DATE
FILMED
6-81
DTIC

AD A099106

AD

TECOM Project No. 7-CO-IL9-TT1-00

USATTC Report No. 790601

MAGNITUDE ESTIMATION: A NEW METHOD FOR MEASURING SUBJECTIVE TEST VARIABLES

By

Dr. Edgar M. Haverland

June 1979

Approved for public release; distribution unlimited.

UNITED STATES ARMY TROPIC TEST CENTER

APO MIAMI 34004

DISPOSITION INSTRUCTIONS*

Destroy this report when no longer needed. Do not return it to the originator.

DISCLAIMER

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents. The use of trade names in this report does not constitute an official endorsement of the use of such commercial hardware or software. This report may not be cited for purposes of advertisement.

* These disposition instructions do not apply to the record copy (AR 340-18).

Final Rept.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TECOM Project No. 7-CO-IL9-TT1-001	2. GOVT ACCESSION NO. A099 206	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Magnitude Estimation: A New Method for Measuring Subjective Test Variables		5. TYPE OF REPORT & PERIOD COVERED In-House Laboratory Research Final Report
		6. PERFORMING ORG. REPORT NUMBER USATTC Report No. 790601
7. AUTHOR(s) Dr. Edgar M. Haverland		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS US Army Tropic Test Center ATTN: STETC-TD APO Miami 34004		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Test and Evaluation Command ATTN: DRSTE-AD-M Aberdeen Proving Ground, MD 21005		12. REPORT DATE June 1979
		13. NUMBER OF PAGES 65
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Human Factors	Subjective Measurement
Magnitude Estimation	Tropic Test Center
Psychophysics	Tropic Testing
Ratio Scaling	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Under the In-House Independent Laboratory Research (ILIR) Program, the US Army Tropic Test Center conducted an investigation of cross-modality matching methods, adapted from those used in studies of the measurement of sensations in the field of psychophysics, for use in measuring subjective variables in human factors evaluations. Magnitude estimation was selected as the desired response mode, and a series of laboratory studies of magnitude estimation of line lengths was carried out. Three field studies were also conducted using magnitude estimation to measure subjective variables important in human factors evaluations.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 (cont)

It was concluded that magnitude estimation is a suitable and practical method for measuring subjective variables in human factors evaluations, and that this method measures these variables better than the usual rating and ranking methods.

TABLE OF CONTENTS

PREFACE

This is the final report of a project carried out at the US Army Tropic Test Center, aimed at improving the measurement of subjective test variables in human factors evaluation. The work was supported by the US Army In-House Laboratory Independent Research Program. US Army Infantry soldiers who participated in the field studies of magnitude estimation were from the 193d Infantry Brigade (Canal Zone). This project was conceived by Dr. D. A. Dobbins, former Chief, Technical Division, USATTC, and preliminary work was done by Roger L. Williamson, USATTC staff. Assistance was given in gathering and analyzing magnitude estimation data in the laboratory studies by Charles M. Kindick, USATTC staff.

SUMMARY

Under the In-House Independent Laboratory Research (ILIR) Program, the US Army Tropic Test Center conducted an investigation of cross-modality matching methods, adapted from those used in studies of the measurement of sensations in the field of psychophysics, for use in measuring subjective variables in human factors evaluations. Magnitude estimation was selected as the desired response mode, and a series of laboratory studies of magnitude estimation of line lengths was carried out. Three field studies were also conducted using magnitude estimation to measure subjective variables important in human factors evaluations. USATTC concluded that magnitude estimation is a suitable and practical method for measuring subjective variables in human factors evaluations, and that this method measures these variables better than the usual rating and ranking methods.

SECTION I. BRIEF OF RESULTS AND CONCLUSIONS

The method of cross-modality matching, as developed in the field of psychophysics for the measurement of sensation, was carefully examined as a possible method for measuring subjective variables in human factors evaluations. Magnitude estimation was selected as the most feasible response mode because of its simplicity, and the fact that everybody is familiar with the number scale and can readily give numbers as estimates of feelings or opinions on a subjective variable.

A series of laboratory experiments with magnitude estimation of line length was carried out in order to see whether the results obtained were in accord with those reported in the literature of psychophysics. It was found that magnitude estimates of line length were very well fitted by a power function and that the exponent of the power function which best fitted the data was approximately .92 to .94. These results were very similar to those reported in the psychophysics literature.

The data on magnitude estimation of line lengths were examined for evidence of the stability (reliability) of measurement, and it was found that stability of measurement, at the level of group means with N = 12, was quite satisfactory. Intraclass correlation coefficients of .94 were obtained.

Three field studies of magnitude estimation were conducted. One involved a comparison of the Personnel Armor System for Ground Troops (PASGT) helmets and vest with the standard helmet and vest with respect to comfort. A second study involved a comparison of four different machine guns with respect to perceived accuracy and ease of operation. In the third study soldiers carried loads ranging from 20 to 50 pounds (9.1 to 22.7 kg) over a 4-kilometer jungle course and then were asked to give magnitude estimates of the difficulty of several parts of the course. The results of the field studies on magnitude estimation agreed with other measures of subjective variables and with objective measures of performance, when these were relevant to the subjective variables being measured.

Magnitude estimation provided more precise measurement of subjective variables than the usual rating and ranking methods, in that it provided measurement on a ratio scale, as compared with the ordinal scale measurement provided by the usual methods for measuring subjective variables. It was also noted that magnitude estimation is a relatively easy and practical method of gathering data on subjective variables.

Further comparisons should be made between magnitude estimation data and those obtained from the usual rating and ranking methods of measuring subjective variables, as well as with relevant objective measures of performance. Also, further empirical and theoretical investigations should be conducted concerning appropriate methods of statistical analysis for magnitude estimation data.

SECTION II. INTRODUCTION

Human factors evaluation of military equipment and materiel involves both objective measures of performance and subjective measures, such as those relating to comfort, preference and confidence. In an effort to gain acceptance and status among their professional peers, human factors specialists have tended to use objective or "hard" measures of performance as much as possible.[1] However, it has always been necessary to use subjective measures in the area of soldier acceptance of equipment and materiel. What a soldier thinks or feels about a piece of equipment is likely to have a strong influence on how effectively he uses that equipment.

Work on a new approach to measuring subjective test data in the US Army test and evaluation setting was begun by Williamson and Dobbins.[2] They completed an extensive review of recent literature and laid out several steps to be accomplished in carrying out the project. This report continues the work on improved methods for measuring subjective test data in the Army test and evaluation setting.

Subjective measures have the reputation of being "soft" in contrast to "hard" measures of performance. Many people are inclined to place less reliance on subjective measures than on objective measures. There are several reasons for the suspicion and uneasiness in regard to subjective measures: (1) subjective measures are likely to be much more variable than objective measures because they are more susceptible to the effects of uncontrolled variables which cannot be anticipated. This means that relatively sophisticated experimental designs and methods of statistical analysis must be used with subjective measures, and some people may have difficulty understanding results presented in these terms; (2) the development of subjective measures requires great effort to communicate clearly to subjects the meaning of the subjective variables on which they are to provide data. Frequently, not enough effort is made with the result that inferior measures are often used and the reputation of subjective measures suffers accordingly; (3) the questionnaire, interview and ranking methods used in gathering subjective test data[3] yield only ordinal data; that is, data which tell us only "more than" or "less than," and not "how much" more than or less than. This last reason involves the quality of measurement, and the objective of the project described in this report is to improve the quality of subjective measurement.

---

1 Klein David, "Social Aspects of Exposure to Highway Crash," <u>Human Factors</u>, pp. 211-219.

2 Williamson, R. L., and Dobbins, D. A. <u>A New Approach Toward Quantifying Subjective Test Data</u>.

3 TECOM Pamphlet 602-1, Vol 1, <u>Man-Materiel Systems Questionnaire and Interview Design (Subjective Testing Techniques)</u>.

Measurement may be thought of as the process of matching the characteristics of objects or entities with a set of categories which constitute a scale. Four different scales of measurement may be distinguished, based on the nature of the categories and the relationships among the categories constituting the measurement scale.[4] The names given to these four scales of measurement are: nominal, ordinal, interval and ratio scales. Quality of measurement depends on the kind of measurement scale used.

Measurement on a nominal scale involves describing objects or entities by sorting them into a set of categories about which one can say only that they differ from each other. This is the lowest form of measurement. Sorting a bowl of mixed fruit into apples, oranges, pears and grapes is an example of measurement on the nominal scale. Classifying persons as male or female is another example.

Measurement on an ordinal scale involves describing objects or entities by sorting them into a set of categories which not only differ from one another, but also have some kind of natural order inherent in the categories. Ranking various fruits on the basis of their sweetness or sourness is an example of ordinal measurement. Assigning grades to students on the basis of the number of correct answers on the final exam is another example.

Measurement on an interval scale involves describing objects or entities by sorting them into a set of categories so that the categories are different from one another, are ordered in some natural manner, and the intervals between adjacent categories constituting the scale are equal. The Fahrenheit and Celsius (centigrade) temperature scales are examples of interval scales, in that equal intervals or units of temperature are measured by equal volumes of expansion. In both cases, arbitrary zero points are designated which do not denote the total absence of heat. Some score scales for achievement tests are interval scales, (those based on percentiles or deciles), if one accepts as legitimate the basis for equalization, which is that the proportion of the population falling in any scaled score interval is equal to that falling in any other numerically equal scaled score interval. Again, zero points on such score scales do not denote a total lack of ability.

Measurement on a ratio scale involves describing objects or entities by sorting them into a set of categories so that the categories are different from one another, are ordered in some natural manner, the intervals between adjacent categories constituting the scale are equal, and one of the categories is a natural zero point for the scale, denoting the total absence of the attribute being measured. The existence of a natural zero point for a scale makes it possible to

---

4 Stevens, S. S., ed., "Mathematics, Measurement and Psychophysics," Handbook of Experimental Psychology, 1951.

form meaningful ratios, and thus to make statements such as "Quantity A is half of quantity B," or "Quantity A is _x_ percent of quantity B." This is the highest form of measurement. The Kelvin temperature scale is an example of a ratio scale, since its zero point (which has never been achieved) corresponds to the complete absence of heat. The basic physical scales, such as length, weight and electrical resistance are also ratio scales.

As stated above, current methods for gathering subjective test data yield only ordinal data. The objective of the project described in this report is to develop methods for measuring subjective test data on a ratio scale, and to investigate the practical problems of using ratio scale measurement in human factors evaluations of subjective variables. It will still be necessary to use suitable experimental designs to control unanticipated variation, and to be quite precise in communicating to soldiers the meaning of the subjective variables on which they are to provide data.

SECTION III. BACKGROUND

Subjective responses have been most carefully and extensively studied in the area of sensation. For well over 100 years, psychologists and physicists have attempted to measure the intensity of sensations, and to relate these measurements to the intensity of the physical stimuli which arouse the sensations. For most of this period, study was concentrated on determining absolute and differential thresholds for various sensory modalities such as vision, hearing, taste and touch. An absolute threshold is the least intense physical stimulus that will reliably arouse a subjective response or a sensation. A differential threshold is the smallest difference between two physical stimuli which can be reliably recognized as producing different subjective responses. The search for an absolute threshold is, of course, a search for a zero point on which to anchor a scale for measuring sensation. And the search for differential thresholds, or "just noticeable differences," is a search for units with which to construct scales for measuring sensations.

As elements to use in fashioning scales for measuring subjective responses, absolute and differential thresholds have not been completely satisfactory. A great many studies of absolute sensory thresholds have shown that subjective responses to weak physical stimuli are shifting and variable, and that there is a zone of uncertainty between a stimulus that is clearly too weak to arouse a subjective response and one that is definitely strong enough to arouse a subjective response. Likewise, studies of differential thresholds have revealed a zone of uncertainty between stimulus differences that are clearly too small to arouse recognizably different subjective responses, and stimulus differences that are definitely large enough to arouse recognizably different subjective responses. Thresholds have been determined, then, by arbitrary statistical methods of dividing these zones of uncertainty and are thus derived from unstable and fluctuating judgments (Stevens, 1951).

During the last 25 years substantial progress has been made toward improving the measurement of sensation (Stevens, 1975).[5] It has been determined that people can easily and confidently make cross-modality matches, such as adjusting the brightness of a light to match the loudness of sounds presented by the experimenter. The sounds should be chosen to cover a substantial part of the range between very faint sounds and very loud sounds. The data from such an experiment consist of the loudness values of the stimulus sounds, expressed in physical terms; and the brightness values, in physical terms, obtained by having the light adjusted by the observer. When the brightness values are plotted against the loudness values on log-log coordinates, the points fall very nearly on a straight line. (An equivalent method is to convert both brightness data and the loudness values to

[5] Stevens, S. S., Psychophysics, 1975.

decibels and plot the points on ordinary linear graph paper.) This result tells us that the subjective brightness (ψ) of the light is a power function of the objective loudness (ϕ) of the sound stimuli: $\psi = \phi^n$. If we take the logarithm of each side of this equation (which is the analytical equivalent of plotting the relationship on log-log coordinates), we obtain: $\log\psi = n\log\phi$. Thus, we see that the exponent n is the slope of the straight line which is obtained when the brightness data are plotted against the loudness values on log-log coordinates.

Over the last 25 years, there has been established an extensive and interwoven set of these power functions which relate various sensory continua to each other. The relationship between any sensory continuum and any other sensory continuum has been found to be a power function, defined by the value of the specific exponent, n, for that particular relationship. Among the sensory continua involved have been: loudness of sound, brightness of light, 60-hertz vibration on the skin, 60-hertz electric current through the fingers* (with a current level high enough to produce sensation but below the levels that would produce pain or "shock"), handgrip force, warmth on the arm, heaviness of lifted weights, pressure on palm of hand, cold on the arm, redness (or saturation) of color, roughness of emery cloth on the skin, length of lines, hardness of rubber balls squeezed,* sweetness, saltiness, sourness, and bitterness of taste, and number or numerosity. The last of these continua, number or numerosity, must be explained further. Data on the relationship of this continuum to any sensory continuum are obtained by presenting observers with stimuli, such as sounds of various loudness, and asking them to produce numbers describing the loudness of the sounds--the louder the sound, the larger the number. This procedure is called magnitude estimation. It is important that the observers <u>not</u> be given any guidance on the scale to be used, such as, "Rate on a scale from 1 to 10 the loudness of these sounds." If observers are given such guidance, they will apportion the provided scale numbers to cover the range of whatever perceptual continuum they are dealing with, and the result will be measurement (at best) on the interval scale (Stevens, 1975, pp. 134-139).

The interwoven set of power functions referred to in the last paragraph exists because relationships between sensory continua have been found to be transitive (Stevens, 1975, pp. 100-107). The results presented in table 1 illustrate this transitivity.

The data in table 1 show that if loudness is matched with vibration, and loudness with shock, and exponents obtained for these power functions, it can be predicted that the exponent obtained when vibration is matched with shock is: $8.46 \div 1.71 = 4.95$, as compared with the experimentally determined exponent of 5.00. Further, if vibration

* Data from these two sensory continua are not so well fitted by a power function as the other continua, for as yet unknown reasons.

Table 1. Experimental Validation of Transitivity

Description of Experiment	Experimentally Determined Exponent	Predicted Exponent
Matching Loudness with Vibration	1.71	(1.69)
Matching Loudness with Shock	8.46	--
Matching Vibration with Shock	5.00	(4.95)

is matched with shock, and loudness with shock, and exponents obtained for these power functions, the exponent obtained can be predicted when loudness is matched to vibration: 8.46 ÷ 5.00 = 1.69, as compared with the experimentally determined exponent of 1.71.

Extensive experimental exploration of transitivities among power function relationships between sensory continua has led to the generalization: from the exponents obtained by experimentally matching any two sensory continua with a third sensory continuum, the exponent for the power function relating two sensory continua can be predicted. Thus the characterization: "interwoven set of power functions."

Because the exponent obtained for a given sensory continuum varies, depending on the sensory continuum against which it is matched (note the two different experimentally determined exponents for loudness in table 1), it is necessary to select a reference continuum. Then exponents for all other sensory continua may be expressed in terms of the reference continuum, which by definition is assigned an exponent of 1.00. The number or numerosity continuum has been widely accepted as the reference continuum. It is convenient because people are almost universally familiar with it and, furthermore, many of the basic measuring scales of physics such as length and mass are linear (exponent = 1.00) against number. Thus, magnitude estimation has become a widely used technique in the measurement of sensation.

The idea has been frequently challenged that legitimate measurement is achieved simply by having people emit numbers in response to physical stimulation of different kinds or intensities. Further, the assertion that this method produces measurement on a ratio scale has been hard for many people to accept. It is true that the necessary elements of a natural zero point and equality of units are not intuitively obvious in magnitude estimation, as they are for basic physical measurements of length and mass; however, laboratory data obtained by the magnitude estimation technique, when subjected to the treatments appropriate for ratio scale data such as the geometric mean and logarithmic transformations, have been very useful in measuring sensation in the field of psychophysics.

Magnitude estimation has also been used on a wide variety of subjective dimensions in other areas of psychology and in political

science, sociology and criminology. Among the successful applications of magnitude estimation have been studies on attitude toward religion, preference for wristwatches, judged quality of handwriting and drawings, esthetic judgments of music, intensity and pleasantness of odors, judgments of masculinity and femininity, a political dissatisfaction scale, judged prestige of occupations, judgments of social status, perceptions of national power, scales of national conflict and cooperation, judged seriousness of crimes generally and of thefts of various amounts of money, and estimates of word frequency (Stevens, 1975, Chapter 8).

SECTION IV. LABORATORY STUDIES OF MAGNITUDE ESTIMATION

At the US Army Tropic Test Center, several laboratory studies of magnitude estimation were carried out before attempting use of the method in field studies. These laboratory studies served as pilot tests to insure that the technique was usable. They actually constituted a calibration step to see if power functions could be obtained with exponents similar to those obtained by other investigators.

As a simple laboratory method of studying subjective variables, magnitude estimation of the lengths of lines was used. The reverse procedure, having subjects draw lines judged to represent (by their lengths) the size of numbers presented to them (line production), was used also.

A. MAGNITUDE ESTIMATION OF LENGTHS OF LINES

1. Experimental Materials and Method: First Experiment

Lines of 1/8, 1 1/8, 2 1/8, 3 1/8, 4 1/8, 5 1/8, 6 1/8 and 7 1/8 inches were drawn, each line on a sheet of 8- by 10 1/2-inch paper. Thirty-six copies of each of these eight sheets of paper were then reproduced to be used as stimuli in this experiment. For each of 12 subjects, three sets of one each of the eight lines of different lengths were selected. Each of these 36 sets of eight sheets of paper was then arranged in random order independently. After each subject had been given instructions, he/she was presented, one at a time, with the 24 sheets of paper which constituted three sets. Thus, each subject was presented with one set of the eight lines of different lengths in random order as a first trial; a second set of the eight lines in a different, independent random order as a second trial; and a third set in a different independent random order as a third trial. For 12 subjects, then, 36 different, independent random orders of presentation of the eight lines of different lengths were prepared.

As a general introduction to both magnitude estimation of line length and line production in response to numbers, the following instructions were given to each subject:

> We're doing some research on subjective rating scales, and we want you to help us. Subjective rating means telling how you feel about something, such as: how good a pair of shoes fits, how comfortable a helmet feels, or how easy (or how difficult) it is to adjust the straps on a pack. It's hard to get very good measurements of this kind of thing and we're trying to improve the methods used in subjective ratings of many kinds of equipment that we test for the Army.

Then the following instructions for magnitude estimation of line length were given to each subject:

You will be presented with a series of lines of various lengths. Your task is to tell how long the lines seem to you by assigning numbers to them. Assign the first line any number that seems appropriate to you. Then assign larger or smaller numbers to the other lines depending on how long they appear to you. You can use any numbers you want: large, small, whole numbers, decimals, or fractions; but please do not use zero or negative numbers. Also, you shouldn't think of the lines as being so many inches or centimeters long. Try to make each number match the length of the line as it appears to you. Please write the number you choose in the space in the lower right hand corner of each page.

The 12 subjects were volunteers from the staff of US Army Tropic Test Center, both men and women, and both military and civilian.

2. Analysis and Results: First Experiment

The magnitude estimates for each trial were arranged in eight columns, one column for each of the various line lengths, and 12 rows for the 12 subjects. Geometric means of the magnitude estimates were computed over the 12 subjects for each of the line lengths for each trial, and for each line length for all three trials combined.*

The lengths of the eight lines were converted to ratios, using the length of the shortest line as the base for the ratios. Similarly, the eight geometric means of the magnitude estimation responses for 12 subjects over all three trials combined were converted to ratios, using the first geometric mean (of the magnitude estimates for the shortest line) as the base for these ratios. Then both of these sets of ratios were converted to decibels by taking the common logarithm of each ratio and multiplying it by 10.**

The logarithms of the ratios could have been used without converting to decibels, but following the conventions established by Stevens (1975), the decibel unit is used here. Using the length of the shortest line and using the geometric mean of magnitude estimates of the shortest line as bases for the conversions to ratios results in the first point being at the origin of the coordinate system, when the points are plotted on decibel scales. Taking ratios and converting them to decibels (or logarithms) makes it possible to plot a power function as

* The geometric mean of $\underline{n}$ numbers is obtained by multiplying all $\underline{n}$ numbers together, and then taking the $\underline{n}$th root of the product.

** Because length of line and number are not obviously analogous either to power or to voltage and current, it was arbitrarily decided to define the decibel scale, for present purposes, as 10 times the common logarithm of the ratio of the lengths of two lines, or of two numbers.

a straight line on linear graph paper. The lengths of the eight lines, the eight geometric means for 12 subjects over all three trials, and the ratios and decibel values obtained from them are shown in table 2.

Table 2. Ratios and Conversions to Decibels: Magnitude Estimates, First Experiment

Stimuli			Response		
Lengths of Lines	Ratios	Decibels	Geom Means of Mag Est	Ratios	Decibels
1/8 in	1	0.00	0.514	1.000	0.00
1 1/8 in	9	9.54	3.956	7.696	8.86
2 1/8 in	17	12.30	6.967	13.554	11.32
3 1/8 in	25	13.98	9.413	18.313	12.63
4 1/8 in	33	15.19	13.073	25.434	14.05
5 1/8 in	41	16.13	17.278	33.615	15.27
6 1/8 in	49	16.90	19.707	38.340	15.84
7 1/8 in	57	17.56	23.380	45.486	16.58

The magnitude estimates are plotted against the lengths of lines in figure 1, using the decibel figures from table 2. The straight line drawn in figure 1 was fitted by least squares to the eight points shown. The slope of this line is .94, which is reasonably close to the value of 1.00 by Stevens (1975).

It can be seen in figure 1 that the eight points lie very nearly on the straight line, as they should if magnitude estimation is a power function of line length.

B. LINE PRODUCTION IN RESPONSE TO NUMBERS

1. Experimental Materials and Method: First Experiment

The numbers 1, 3, 10, 30, 100, 315, 1,000, 3,150 and 10,000 were chosen as stimuli on the basis that their logarithms are (approximately) evenly distributed over the range, 0 to 4: 0, 0.5, 1.0, 1.5, . . ., 4.0. Each of these nine numbers was written on an 8- by 10 1/2- inch sheet of paper, and 36 copies were reproduced of each of these nine sheets of paper. In the same manner as was done for the experiment on magnitude estimation of the length of lines, three sets of one each of the nine numbers were selected for each of the 12 subjects. Each of these 36 sets of nine sheets of paper was then arranged in random order independently. After each subject had been given instructions; he/she was presented, one at a time, with the 27 sheets of paper (numbers) which constituted three sets. Thus, each subject was presented with one set of the nine numbers in random order as a first trial; a second set of the nine numbers in a different,

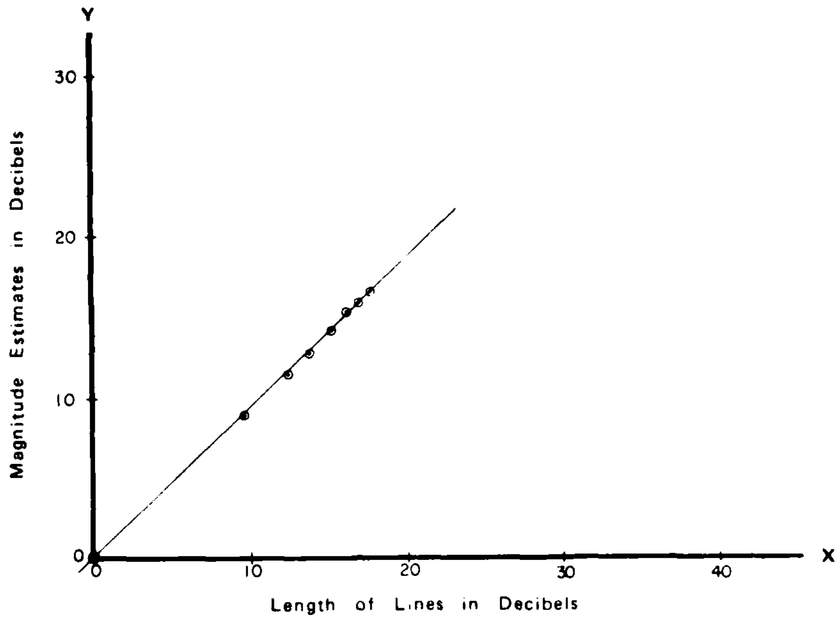

Figure 1. Magnitude Estimation of Lengths of Lines. First Experiment: Twelve Subjects, Three Trials on Each of Eight Line Lengths. Best-fitting Line (least squares): Y = -.09 + .94X.

independent random order as a second trial; and a third set in a different, independent random order as a third trial. For 12 subjects, then, 36 different, independent random orders of the nine numbers were prepared.

The subjects were given the same "general introduction" instructions as were quoted in the last section on magnitude estimation, and then were given the following specific instructions for the line production experiment:

> You will be presented with numbers ranging from 1 to 10,000. Your task is to draw a line for each number so that the length of the line represents the size of the number. Draw the line from left to right across the page. Make the line as long as you think the number is large. Try not to think of inches or centimeters, or any other units of length.

The subjects were given a plastic straight-edge, which was not marked with a scale of any kind. The 12 subjects were volunteers from the staff of US Army Tropic Test Center; both men and women, and both military and civilian. None of this group of 12 subjects were persons who had been subjects for the first magnitude estimation experiment.

2. Analysis and Results: First Experiment

The lines drawn by the subjects in response to the numbers were measured and their lengths (in millimeters) recorded in nine columns and 12 rows in the same manner as was done with the magnitude estimation. Geometric means of the line lengths were computed over the 12 subjects for each of the nine stimulus numbers for each trial and for each number for all three trials combined.

The nine numbers used as stimuli were converted to ratios and then to decibels. Similarly, the geometric means of the line lengths for all three trials combined were converted to ratios, and then to decibels, as was done in the magnitude estimation experiment. The nine numbers, the nine geometric means for 12 subjects over all three trials, and the ratios and decibel values obtained from them are shown in table 3.

Table 3. Ratios and Conversion to Decibels: Line Production, First Experiment

Stimuli			Responses		
Numbers	Ratios	Decibels	Geom Means of Line Lengths	Ratios	Decibels
1	1	0.00	1.133 in	1.000	0.00
3	3	4.77	2.868 in	2.531	4.03
10	10	10.00	5.160 in	4.554	6.58
30	30	14.77	8.458 in	7.465	8.73
100	100	20.00	13.445 in	11.867	10.74
315	315	24.98	22.983 in	20.285	13.07
1,000	1,000	30.00	48.245 in	42.582	16.29
3,150	3,150	34.98	94.253 in	83.189	19.20
10,000	10,000	40.00	225.982 in	199.455	23.00

The geometric means of the line lengths are plotted against the numbers, both being expressed in decibels, in figure 2. The straight line drawn in figure 2 was fitted by least squares to the nine points shown. It can be seen that the nine points lie quite close to the line as they should if line production is a power function of the numbers used as stimuli. However, the slope of this line is only .53, which is considerably less than the .94 slope obtained when magnitude estimates were plotted against line lengths. This is an example of the regression effect which may be described as a tendency of subjects to restrict the range of the variable they control, i.e., numbers produced in the case of magnitude estimation and line length in the case of line production (Stevens, 1975, pp. 271-281). However, the difference in slopes (exponents) seems rather large. For this reason, further experiments were undertaken, first with line production in response to numbers, since it was the exponent obtained in this experiment which seemed to differ so much from the expected value.

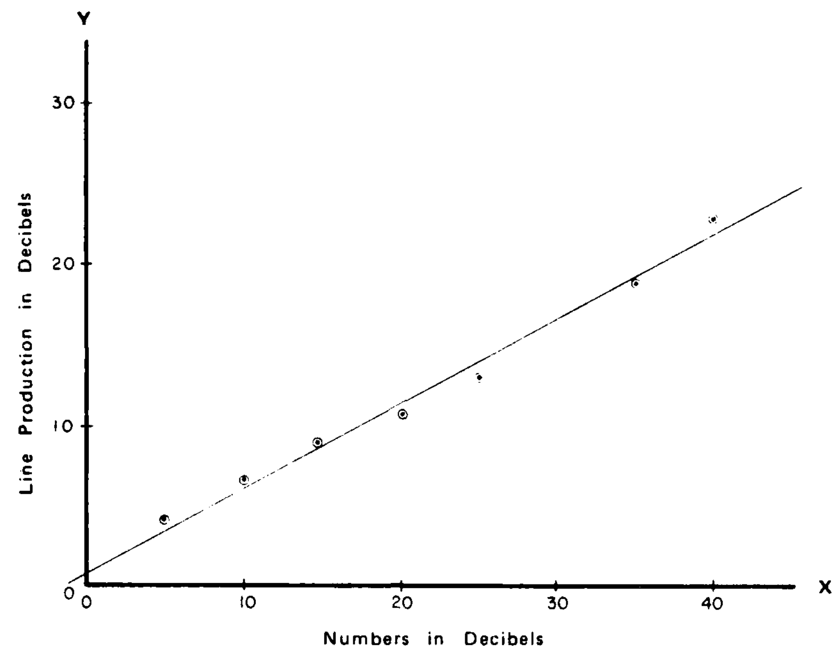

Figure 2. Line Production in Response to Numbers. First Experiment: Twelve Subjects, Three Trials on Each of Nine Numbers. Best-fitting Line (least squares): $Y = .61 + .53X$.

3. Experimental Materials and Method: Second Experiment

The range of numbers used as stimuli in the first experiment was quite great (1 to 10,000, or 40 decibels), but the subjects were limited in the length of line they could draw to about 267 millimeters by the 10 1/2-inch length of the sheet of paper. Therefore, it was hypothesized that this restriction had prevented the subjects from varying the lengths of lines drawn over a range comparable to the range of numbers used as stimuli. The line production experiment was repeated, using pieces of paper 14 15/16 inches long, which permitted the subjects to draw lines as long as 380 millimeters. The same numbers were used as stimuli as in the first experiment. Again, subjects were 12 volunteers from the US Army Tropic Test Center staff, different persons from those who served as subjects for the first experiment.

4. Analysis and Results: Second Experiment

When the results of the second experiment were analyzed and plotted in the same fashion as those of the first experiment in line production, the resulting plot appeared as shown in figure 3. Again, the points fall very nearly on a straight line, but the slope of this line is only .50. Therefore, the hypothesis, that restricted space for line drawing lowered the exponent of the power function relating

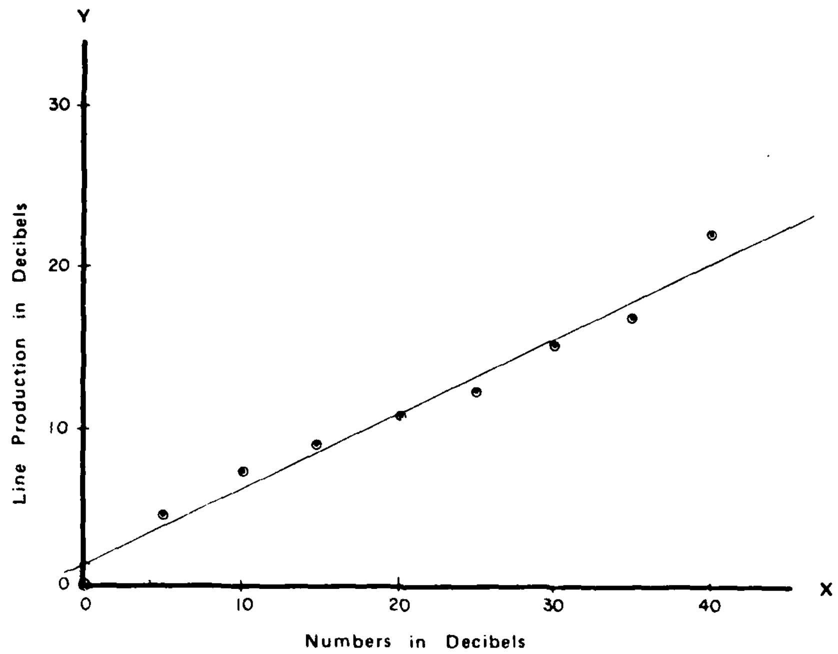

Figure 3. Line Production in Response to Numbers. Second Experiment: Twelve Subjects, Three Trials on Each of Nine Numbers. Best-fitting Line (least squares): Y = 1.01 + .50X.

line production to numbers, was rejected and a third experiment was performed.

5. Experimental Materials and Method: Third Experiment

In the second experiment, the ratio of the largest number used as a stimulus to the smallest number was 10,000::1, thus yielding a 40-decibel range on the number scale in figure 3. But, the ratio of the geometric mean of line lengths produced in response to the largest number, to the geometric mean of line lengths produced in response to the smallest number was only 169.88::1, yielding only a 22.30-decibel range on the line production scale in figure 3. Therefore, in the third experiment an attempt was made to reduce this difference between the two ratios by using as stimuli the set of numbers: 1, 2, 4, 8, 16, 32, 64, 128, and 256; which has a ratio between largest and smallest of only 256::1, yielding a 24.08-decibel range on the number scale. The experiment was administered in the same fashion as the first experiment on line production, with the exception that the instructions to the subjects were modified to specify a range of numbers from 1 to 256. The 13 subjects were volunteers from the US Army Tropic Test Center staff; men and women, military and civilian.

6. Analysis and Results: Third Experiment

The results of the third experiment were analyzed and plotted in the same fashion as those of the first experiment in line production, producing the plot shown in figure 4. The points fall very nearly on a straight line, and the slope of the line this time is .71, which is considerably closer to the value of .94 obtained from magnitude estimation in response to line length. The decibel value for the geometric mean of the lengths of lines drawn in response to the largest number declined from 22.30 in the second experiment (in which the decibel value for the largest number was 40.00) to 18.42 in the third experiment (in which the decibel value for the largest number was 24.08). Thus, the decrease in the range of decibel values for line production between the two experiments was relatively less than the decrease in range of decibel values produced by reducing the range of stimulus numbers to 256::1, and an increased slope for the best fitting line results.

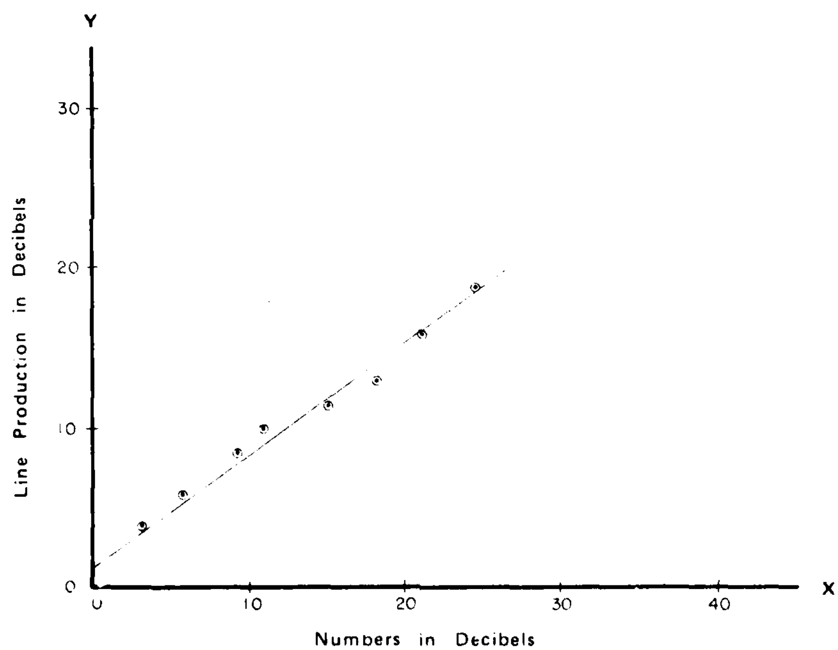

Figure 4. Line Production in Response to Numbers. Third Experiment: Thirteen Subjects, Three Trials on Each of Nine Numbers. Best-fitting Line (least squares): Y = 1.05 + .71X.

C. CONFIRMATORY EXPERIMENTS

At this point in the investigation, two confirmatory experiments were performed, one with magnitude estimation and one with line production.

1. Magnitude Estimation: Experimental Materials and Method

As stimuli for this experiment, 10 lines of .12 (1/8), .22, .40, .70, 1.25, 2.22, 3.95, 7 Q3 12.50 and 22.23 inches in length were chosen. The basis for choosing lines of these particular lengths was that when they are converted to decibels using .125 (1/8) inch as the base for ratios, a series of equally spaced numbers would be obtained: 0, 2.5, 5.0, . . ., 20.0, 22.5 decibels. Also, the range on the length of lines variable is extended to 22.5 decibels, compared with 17.6 decibels in the first experiment on magnitude estimation.

Ten pieces of 6- by 24-inch poster board were selected, and lines, one line per board, of the lengths described above were drawn with a pen which produced a line approximately 1 millimeter wide. On the back of each piece of poster board a number was written: 1 for the board on which the .125-inch line was drawn, 2 for the board on which the .22-inch line was drawn, . . ., and 10 for the board on which the 22.23-inch line was drawn.

Thirty-six independent, random orders of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 were prepared from random number tables. This provided for three trials for each of 12 subjects. The subjects were given the same "general introduction" instructions as in the earlier laboratory experiments, followed by the instructions for magnitude estimation of line length and modified to request that the subjects write their magnitude estimations on a data sheet, rather than "in the lower right hand corner of each page." The 12 subjects were volunteers from the staff of US Army Tropic Test Center, both military and civilian. By coincidence all 12 subjects were male.

In working with each subject, the experimenter first arranged the 10 pieces of poster board in the random order for the first trial for that subject, then showed the subject the lines, one at a time, with instructions to write his magnitude estimation of the length of the line on a data sheet before showing him the next line. After all 10 lines had been shown to the subject on the first trial, the experimenter arranged the 10 pieces of poster board in the order for the second trial for that subject, and proceeded to show them to the subject in the same manner as had been done on the first trial. The procedures for the third trial were the same as for the first two trials.

2. Magnitude Estimation: Analysis and Results

The magnitude estimates were arranged in 10 columns, one for each of the stimulus lines, with estimations for 12 subjects in each column. Geometric means of the magnitude estimates were computed over the 12 subjects for each of the 10 line lengths for each trial, and for each line length, for all three trials combined, as had been done in the first magnitude estimation experiment. The lengths of the 10 lines and the 10 geometric means of the magnitude estimates were converted to ratios and then to decibels, in the same manner as in the

first magnitude estimation experiment. The magnitude estimates were then plotted (in decibels) against the lengths of the stimulus lines (also in decibels) in figure 5.

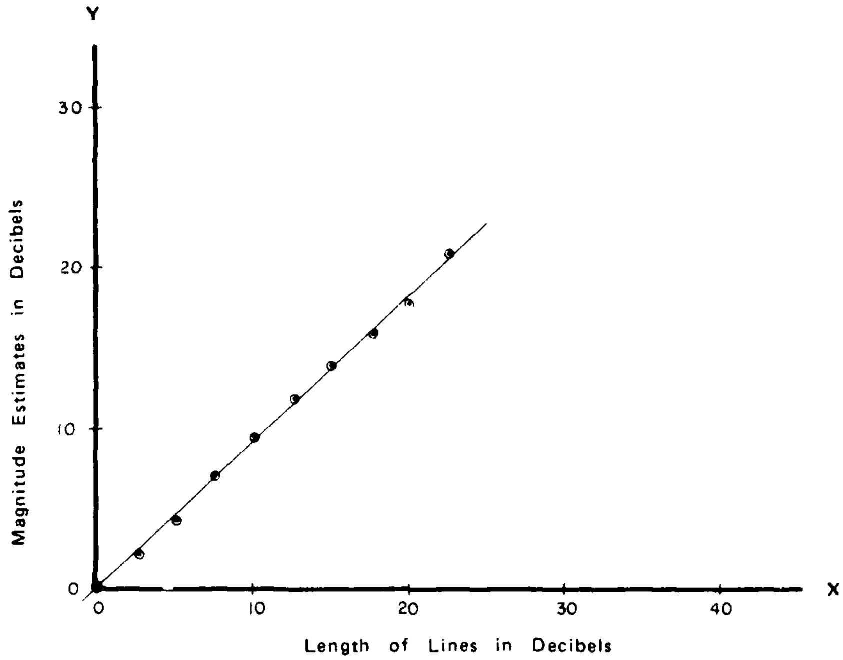

Figure 5. Magnitude Estimation of Lengths of Lines. Confirmatory Experiment: Twelve Subjects, Three Trials on Each of 10 Line Lengths. Best-fitting Line (least squares): $Y = -.11 + .92X$.

The slope of the best-fitting line in figure 5 is .92, which agrees quite well with the slope of .94 obtained in the first magnitude estimation experiment.

3. Line Production: Experimental Materials and Method

A comparison between the second and third line production experiments, previously described, shows that reducing the range of the set of numbers used as stimuli was apparently a step in the right direction toward the objective of obtaining a power function with an exponent nearer to 1.00 (see discussion in Section IV, B, 6, Analysis and Results: Third Experiment). Therefore, in this confirmatory experiment on line production in response to numbers, a further step in this direction was taken by reducing the range of the set of numbers even more so that the ratio of the largest number to the smallest was 63::1. The following 10 numbers were used as stimuli: 1, 1.6, 2.5, 4, 6.3, 10, 16, 25, 40, and 63. The spacing within this set of numbers

was chosen to produce decibel values at approximately equal intervals: 0, 2, 4, . . ., 18. This range, from 0 to 18 decibels for the set of numbers used as stimuli in this experiment, thus approximates the 0 to 18.42 decibels range for length of lines drawn in response to numbers in the third line production experiment previously described.

Use of the longer paper (allowing lines as long as 380 millimeters to be drawn) was continued from the second and third line production experiments previously described. The experimental materials were prepared in a fashion analogous to the previous line production experiments. Each of the 10 numbers was written on three sheets of paper for 12 subjects, and the sheets of paper were then sorted into 36 sets of 10 sheets, so that for each subject three sets of the 10 numbers were available for trials one, two and three. Each of these 36 sets was then arranged in a different, independent random order. The experiment was administered in the same fashion as the earlier line production experiments, with the exception that the instructions to the subjects were again modified, this time to specify a range of numbers from 1 to 63. As before, the subjects were volunteers from the US Army Tropic Test Center staff, both military and civilian. None participated in the confirmatory experiment on magnitude estimation, though some had participated in one or both types of experiments previously carried out in this project.

4. Line Production: Analysis and Results

The results of this confirmatory experiment on line production in response to numbers were analyzed and plotted in the same fashion as those of earlier line production experiments. Figure 6 shows the plotted results. Again the points fall very nearly on a straight line, and the slope of the best-fitting line is .78, which is somewhat nearer 1.00 than the slope of .71 obtained from the third line production experiment previously described. When the slope of .78 obtained in this confirmatory line production experiment is compared with the slope of .92 obtained in the confirmatory magnitude estimation experiment, it can be seen that the regression effect (Stevens, 1975, pp. 271-281) is still present, but is much reduced from the corresponding comparison of .53 with .94, obtained in the first experiments in this series.

D. Stability of Magnitude Estimation Data

An answer to the question of how stable magnitude estimation data are may be obtained by examining the results of the three trials separately. Because our use of the data yielded by magnitude estimation will be different from the uses often made of data obtained from responses made to psychological tests and rating instruments, our evaluation of the stability of magnitude estimation data will be different from the usual evaluation of stability (or reliability) of the data yielded by psychological tests and rating instruments.

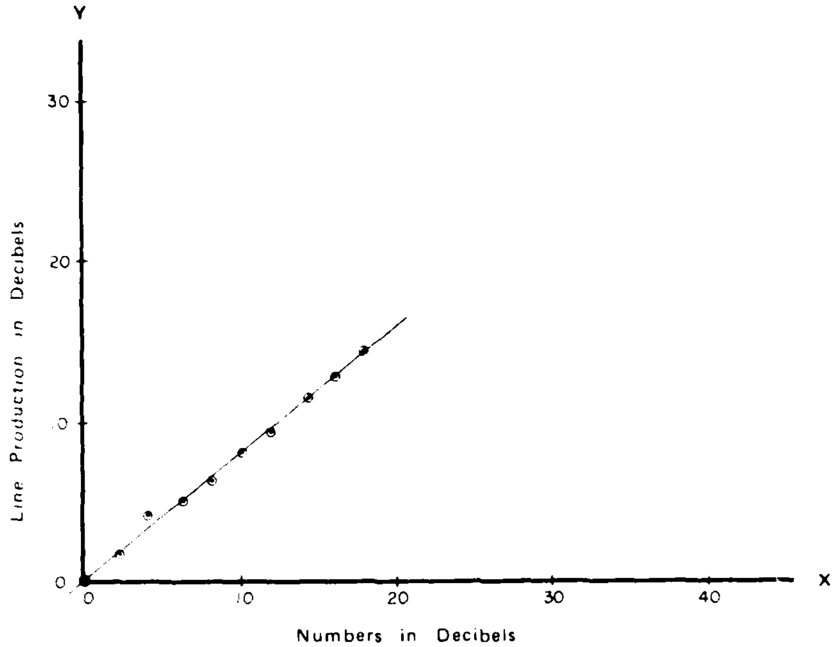

Figure 6. Line Production in Response to Numbers. Confirmatory Experiment: Twelve Subjects, Three Trials on Each of 10 Numbers. Best-fitting Line (least squares): Y = .17 + .78X.

Data from psychological tests and rating instruments are often used to help make decisions about individual persons whose responses constitute the data. In this case, therefore, it is important that there be reasonable stability in the data at the level of the individual person. Magnitude estimation is being proposed here as a technique for providing data to aid in decisions, not about the individual persons whose responses constitute the data, but rather in decisions concerning the items of materiel which the persons are evaluating in some way. It is assumed that when decisions are being made concerning materiel being evaluated, by means of the magnitude estimation technique, data will be available from groups of at least 10 to 12 persons. Therefore, the stability of the data yielded by magnitude estimation at the level of groups of 10 or 12 persons will be evaluated.

1. Comparison of Geometric Means for the Three Trials

One method of examining the stability of magnitude estimation data at the level of 12-person groups is to compare the geometric means of the magnitude estimates for each line length for each of the three trials. Table 4 presents the geometric means of the magnitude estimates for each trial separately, as well as for all three trials combined, for the first experiment in magnitude estimation.

Table 4. Geometric Means of Magnitude Estimates of Line Lengths: First Experiment

Lengths of Lines	1st Trial	2nd Trial	3rd Trial	3 Trials Combined
1/8 in	0.450	0.630	0.479	0.514
1 1/8 in	3.743	3.913	4.227	3.956
2 1/8 in	7.088	7.374	6.470	6.967
3 1/8 in	8.696	10.094	9.503	9.413
4 1/8 in	13.111	13.250	12.862	13.073
5 1/8 in	15.529	16.191	20.215	17.278
6 1/8 in	16.594	21.359	21.594	19.707
7 1/8 in	19.218	26.018	25.561	23.380

Comparing geometric means across the rows of table 4 shows fairly good stability, though on the first trial there appears to have been some inhibition against giving larger magnitude estimates in response to the longest lines, in contrast to trials 2 and 3. The geometric means of the magnitude estimates for each trial separately, and for all three trials combined, for the confirmatory experiment in magnitude estimation are presented in table 5.

When the geometric means in each row of table 5 are compared, considerable stability is apparent. The first-trial reluctance of subjects to give larger magnitude estimates in response to the longer lines, so apparent in the first experiment, was not found in this confirmatory experiment.

Table 5. Geometric Means of Magnitude Estimates of Line Lengths: Confirmatory Experiment

Lengths of Lines	1st Trial	2nd Trial	3rd Trial	3 Trials Combined
0.125 in	0.199	0.167	0.253	0.204
0.22 in	0.299	0.335	0.352	0.328
0.40 in	0.480	0.435	0.702	0.527
0.70 in	1.034	0.968	0.932	0.977
1.25 in	1.623	1.675	2.009	1.761
2.22 in	2.668	3.097	3.445	3.053
3.95 in	4.689	4.774	5.243	4.896
7.03 in	7.865	6.995	7.753	7.527
12.50 in	13.197	10.899	13.145	12.365
22.23 in	23.857	24.907	24.737	24.496

2. Plot of Decibel Values for the Three Trials

Another method of examining the stability of magnitude estimation data is to plot the decibel values of the magnitude estimates for the three trials separately, in a fashion analogous to figure 1.* Figure 7 shows such a plot for the first experiment in magnitude estimation.

The line in figure 7 is the best fitting line for the points based on the geometric means of all three trials combined, the same line as appears in figure 1. Though there is some scatter about this line, the points for the three trials for a line 1 1/8 inches (9.54 decibels) long do not overlap with those for a line 2 1/8 inches (12.30 decibels) long, etc., until reaching the three lines of greatest length, where the intervals on the decibel scale between lengths of lines become quite small. The inhibition against giving larger magnitude estimates in response to the longest lines on the first trial, noted above, shows up clearly in figure 7.

Figure 8 shows a plot of the decibel values of the magnitude estimates for the three trials separately, in the confirmatory experiment. Here, where the intervals on the decibel scale for length of lines are equal, there is no overlap between the points for the three trials with any line length and those for the three trials with any adjacent line length.

3. Best-Fitting Lines for Trials 1, 2 and 3

A third method for examining the stability of magnitude estimation is to compute the best-fitting lines for the points of trials 1, 2, and 3. This was done, and the slopes for these three best-fitting lines are .93, .90, and .98, respectively, for the first experiment in magnitude estimation. These slopes may be compared with a slope of .94 for the best-fitting line for the points of all three trials combined. In the confirmatory experiment the slopes of the best-fitting lines for trials 1, 2 and 3, respectively, are .93, .93 and .89. These slopes may be compared with a slope of .92 for the best-fitting line for the points of all three trials combined in the confirmatory experiment.

* The ratios of the geometric means of the magnitude estimates for the three trials taken separately are taken to the same base, .514, which is the geometric mean of the magnitude estimates in response to the shortest line for all three trials taken together (table 2). Increasing or decreasing the base to which these ratios are taken simply lowers or raises the points on the graph. Therefore, using the same base for the ratios for the three trials taken separately provides a common reference framework for the three sets of points.

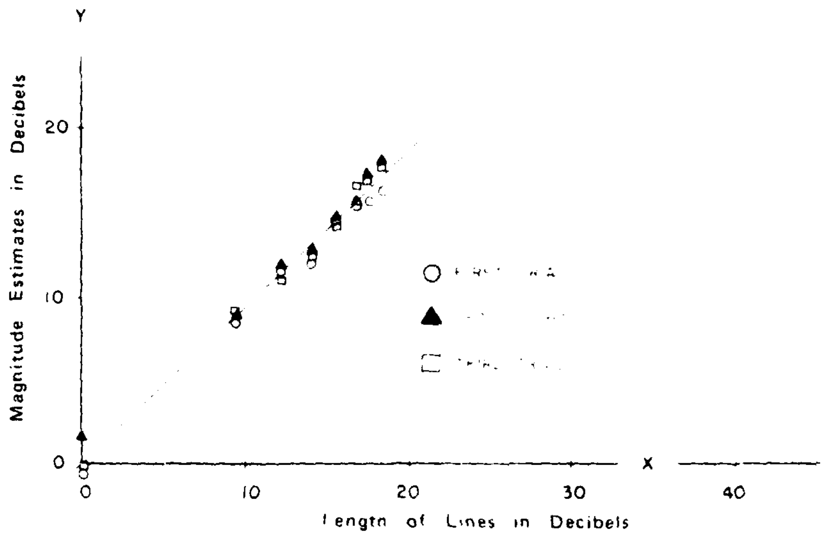

Figure 7. Stability of Magnitude Estimation Data. First Experiment: Twelve Subjects, Three Trials on Each of Eight Line Lengths Best-fitting Line (least squares): Y = -.09 + .94 X, Based on Geometric Means of All Three Trials Combined (figure 1).

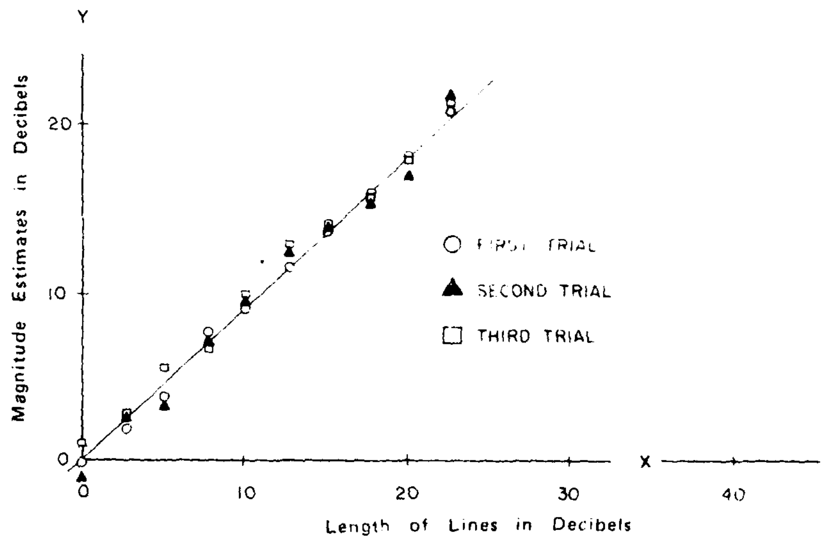

Figure 8. Stability of Magnitude Estimation Data. Confirmatory Ex- iment: Twelve Subjects, Three Trials on Each of 10 Line Lengths. Best-fitting Line (least squares): Y = -.11 + .92 X, Based on Geometric Means of All Three Trials Combined (figure 5).

4. Stability of Magnitude Estimation Data as Measured by the Intraclass Correlation Coefficient

The intraclass correlation coefficient involves an analysis of variance approach to the stability (reliability) of measurements.[6] The mathematical model appropriate for this study of magnitude estimation data is that involved in Case 2, as described in the recent Shrout and Fleiss article on intraclass correlations.[7] In this study we have a sample of persons, each of whom has rated (made magnitude estimates in response to) each of a number of lines of different lengths. We desire to generalize from this sample of persons (raters or judges) to a population of persons; therefore, the person or subject variable in the analysis of variance is a random effect. Further, as was discussed at the beginning of this section, we are interested in the stability of the mean of magnitude estimates made by a group of 10 to 12 persons, rather than in the stability of a magnitude estimate made by one person. Therefore, the formula used to compute intraclass correlation coefficients in this study was that appropriate for mean ratings of a group of raters (Shrout and Fleiss, p. 426).

This intraclass correlation coefficient may be thought of as the ratio of the component of variance due to treatments (in this case, the individual stimulus lines of different lengths) to the sum of the components of variance due to treatments, subjects, interaction between treatments and subjects, and error. In other words, this intraclass correlation coefficient tells us the proportion of the total variance that is accounted for by the treatments.

The intraclass correlation coefficients for the two magnitude estimation experiments are presented in table 6. It can be seen that the coefficients in the first experiment are notably lower than those in the confirmatory experiment, but that they increase from the first, to the second, to the third trial. This means that the variability between subjects in magnitude estimations was much greater in the first experiment than in the confirmatory experiment, and that this variability decreased from trial to trial in the first experiment. The first magnitude estimation experiment was done at the very beginning of this series of laboratory experiments. No ready explanation has been developed for this trend towards increasing stability in the first experiment. It is apparent from the intraclass correlation coefficients obtained in the confirmatory experiment, however, that highly stable (reliable) magnitude estimation data can be obtained.

---

6 Winer, B. J. Statistical Principles in Experimental Design, pp. 283-296.

7 Shrout, P. E., and Fleiss, J. L. "Intraclass Correlations: Uses in Assessing Rater Reliability," Psychological Bulletin, pp. 420-428.

Table 6. Intraclass Correlation Coefficients

	Trials 1	2	3
First Experiment	.57	.61	.67
Confirmatory Experiment	.94	.94	.94

Since the intraclass correlation coefficient is a ratio of variance accounted for by the treatments to total variance, these obtained intraclass correlation coefficients of .94 are equivalent to reliability coefficients of the usual kind of .97 (= the square root of .94).

Stability of line production data was found to be approximately the same as that of the magnitude estimation data. However, it is not presented and discussed in this report, because the magnitude estimation response mode has been chosen as the most feasible and convenient response mode for use in human factors evaluations.

SECTION V. FIELD STUDIES OF MAGNITUDE ESTIMATION

After some experience with laboratory studies of magnitude estimation, field studies were undertaken. Three different field studies were performed: (1) a comparison of Personnel Armor System for Ground Troops (PASGT) helmets and vests with standard helmets and vests with respect to comfort; (2) a comparison of four different machine guns with respect to perceived accuracy, ease of opening and several other features of the weapons; and (3) a study in which soldiers carried loads ranging from 20 to 50 pounds over a 4-kilometer course in the jungle and were asked to give magnitude estimates of the difficulties of various parts of the course.

In field studies of magnitude estimation, it is usually not possible to control and measure stimulus variables. This means that clearcut relationships between physical stimulus variables and subjective response measures cannot be shown, as was done in the laboratory experiments. In addition, the subjective variables measured with the magnitude estimation technique in field studies are likely to be complex functions of a number of physical stimulus variables acting together. Thus, perceived difficulties of parts of the 4-kilometer jungle course are likely to depend not only on the loads carried by the soldiers, but also on temperature and humidity, whether it is raining or not, the physical condition of individual soldiers, and a host of conditions internal to individual soldiers which may be lumped together under a label such as "morale" or "motivation."

Having demonstrated that the magnitude estimation technique does a good job of measuring subjective variables in laboratory experiments, where the physical stimulus can be precisely controlled and where a close relationship can be shown between the physical stimulus and a subjective variable, the next step is to use magnitude estimation in field studies involving human factors evaluations. In these field studies, where stimulus variables cannot be precisely controlled and subjective variables are likely to depend on a number of stimulus variables acting together in a complex fashion, it will probably not be possible to demonstrate close relationships between objective stimulus variables and subjective variables, as was done in the laboratory experiments. Rather, it will be assumed that magnitude estimation, when used in the less controlled and defined setting of field experiments, will continue to do a competent job of measuring subjective variables, as it did in the laboratory experiments.

A. COMPARISON OF HELMETS AND VESTS

1. Data Acquisition Procedures

Magnitude estimation data were gathered during the Development Test II of the Personnel Armor System for Ground Troops (PASGT), which

the US Army Tropic Test Center carried out in 1976-1977.[8] In this test, soldiers traversed a 4-kilometer jungle course (known as the Man-Pack Portability Course (MPPC), described in detail in reference)[9] repeatedly over a period of several days, each time wearing a different helmet-vest combination. The soldiers also performed on a laser-rifle range and a land navigation course in the jungle each day before they traveled the MPPC and again after they traveled the MPPC. At the end of the test period, when the soldiers had worn all of the helmet-vest combinations, 20 of them were asked to make magnitude estimations of the comfort of the helmets and vests.

The following helmet-vest combinations were worn by the soldiers:

a. Kevlar helmet (38 oz/ft^2) (PASGT-1) with Kevlar vest.
b. Kevlar helmet (30 oz/ft^2) (PASGT-2) with Kevlar vest.
c. Standard M-1 helmet with standard B-nylon vest.

The following instructions were given to the soldiers as the magnitude estimate data were gathered:

> We're trying out a new way of asking you what you think of this equipment. Think about the overall comfort of the two different vests you wore. Let's say that a very large number represents the most comfortable vest you can think of, and a very small number represents the most uncomfortable. Now think of a number that represents how comfortable or uncomfortable you think the standard vest was. Please write this number in the blank space next to "Standard Vest."
>
> Now think of another number that represents how comfortable or uncomfortable you thought the new vest was. Please write this number in the blank space next to "New Vest." Remember, if you thought the new vest was more comfortable than the standard vest, you should pick a larger number than you did for the standard vest. If you thought the new vest was less comfortable than the standard vest, you should pick a smaller number than you did for the standard vest. The bigger the difference in comfort, the bigger the difference between the numbers you pick should be.
>
> Now, let's think about the comfort of the three different helmets you wore. Please write numbers in the spaces next to "PASGT-1 Helmet," "PASGT-2 Helmet," and "Standard Helmet," to show how comfortable you felt each helmet was. Remember, the

---

8 Haverland, E. M.; Novak, C. A.; Johnson, R. L., Jr.; Williamson, R. L.; and Kindick, C. M. Development Test II of Personnel Armor System for Ground Troops (PASGT).

9 Test Operations Procedure (TOP) 1-3-550, Man-Pack Portability Testing in the Tropics.

more comfortable you felt the helmet to be, the larger the number; and the less comfortable, the smaller the number.

2. Results

The arithmetic mean of the magnitude estimates of the comfort of the standard vest was 2.15, while that for the new vest (PASGT) was 53.00.* This difference seems quite large, but the between-subjects variability was also large, since some subjects restricted their magnitude estimates to as little as two points, while others let their magnitude estimates range over more than 200 points. Nevertheless, a correlated t-test for difference between the magnitude estimates for standard and new vests yielded a value for t of 3.54. With 19 degrees of freedom and a two-tailed test, the probability of obtaining a value of t this large, if there were no difference in magnitude estimates of comfort for the two vests, is less than .01.

The arithmetic means of the magnitude estimates of the comfort of the three helmets are shown below:

PASGT-1 Helmet	50.90
PASGT-2 Helmet	57.55
Standard Helmet	2.25

A repeated measures analysis of variance of the magnitude estimates of the comfort of the three helmets is given in table 7.

Table 7. Analysis of Variance of Magnitude Estimates for PASGT-1, PASGT-2 and Standard Helmets

Source Variance	SS	df	MS	F	p
Between subjects	71,572.73	19	--	--	--
Within subjects	99,648.67	40	--	--	--
Helmets	36,460.90	2	18,230.45	10.96	<.001
Error	63,187.77	38	1,662.84	--	--
TOTAL	171,221.40	59			

* Though magnitude estimation yields measurement on a ratio scale, special statistical techniques for determining the significance of differences between geometric means are not available. Therefore, the usual statistical techniques (t-tests and analysis of variance) which are appropriate for determining differences between arithmetic means are used in this report.

3. <u>Discussion</u>

These results--the new vest (PASGT) being rated more comfortable than the standard vest, and both of the two new PASGT helmets being rated more comfortable than the standard helmet--agree with the results of a large number of both objective performance test results and subjective questionnaire results obtained during the PASGT test and documented in the test report (Haverland, <u>et. al</u>.). The results obtained with magnitude estimation appear more clear-cut than do the results of the performance tests and questionnaires. Of course, the performance tests and questionnaires covered a much wider variety of variables than the magnitude estimates of comfort, and for this reason should be depended upon as giving much more comprehensive evidence of the superiority of the PASGT equipment to the standard vest and helmet, than the magnitude estimates of comfort. Nevertheless, magnitude estimation appears to have done a good job of measuring differences in comfort between the new PASGT equipment and the old standard vest and helmet.

B. COMPARISON OF FOUR MACHINE GUNS

1. <u>Data Acquisition Procedures</u>

Magnitude estimation data were gathered near the end of the Machine Gun Accuracy and Dispersion (MAD) test conducted at USATTC during 1977-1978. In this test nine different machine guns were fired at 32-foot square targets (at ranges of 300 and 600 meters) by regular troops (MOS 11B, Infantryman, M60 machine gun qualified) from the 193d Infantry Brigade (Canal Zone), and large amounts of data on miss-distances were gathered. The 10 soldiers who provided the magnitude estimation data had fired these particular four machine guns over a period of 8 weeks.

The four machine guns were:

a. MG-1A3, a German 7.62-mm weapon with a heavy barrel. The weapon weighed approximately 35 pounds.

b. RPK, a Soviet 7.62-mm weapon--a member of the AK47 family. It is a light weapon, weighing approximately 15 pounds.

c. PKM, a Rumanian 7.62-mm weapon. It is of intermediate weight, approximately 23 pounds.

d. M60, the standard US Army machine gun (7.62 mm). It is also of intermediate weight, weighing approximately 23 pounds. The soldiers had had more experience with this weapon than with the other three, having fired it extensively before they participated in this test.

The soldiers were given the following instructions:

Number Rating of the MAD Weapons

You have fired four different weapons during the time you have been helping us on this project: the MG-1A3 (German 7.62 mm), RPK (Soviet), PKM (Rumanian), and M60 machine gun. Now, we're trying out a new way of asking you what you think of these weapons. To start with, think about the accuracy of the weapons you've fired. Let's say that a very small number represents a weapon that was very inaccurate, and a very large number represents a weapon that was extremely accurate.

Now think of a number that represents how accurate or inaccurate you thought the MG-1A3 was. You should not use zero or negative numbers. Please write this number in the blank space on the first line below MG-1A3.

Now think of another number that represents how accurate or inaccurate you thought the RPK was. Please write this number in the blank spaces on the first line below RPK. Remember, if you thought the RPK was more accurate than the MG-1A3, you should pick a larger number than you did for the MG-1A3. If you thought the RPK was less accurate than the MG-1A3, you should pick a smaller number than you did for the MG-1A3. The bigger the difference in accuracy, the bigger the difference between the numbers you pick for the two weapons should be.

Now, go ahead and pick numbers to represent the accuracy or inaccuracy of the PKM and the M60 and write them in the spaces on the first line under PKM and M60. The more accurate you felt the weapon was, the larger the number you should choose for it. Each of you should choose your numbers by yourself, without talking to anybody else about it. You can talk about it after we're finished.

Wait until everybody has finished the first line on accuracy.

Now think about how easy or hard it was to open the weapons. If you thought it was easy to open a weapon, you should give that weapon a large number--the easier it was to open the weapon, the larger its number should be. If you thought it hard to open a weapon, you should give that weapon a small number--the harder it was to open, the smaller its number should be. Again, you should not use zero or negative numbers. Now let's go ahead and put down numbers for how easy or hard you thought it was to open the four weapons.

Are there any questions? Now let's go ahead and put down numbers for the other six things I'm asking you about these

weapons. Think about each question for a minute or two, and then put down a number for each of the four weapons.

The data sheet on which the soldiers wrote their magnitude estimates is reproduced below, with means of obtained data entered in response spaces (table 8).

The instructions were given to the 10 soldiers in a group, and they went ahead with making their magnitude estimates. As the instructions were given, one soldier asked if he should use a scale from 1 to 10. The experimenter explained that they should use any numbers they wanted, other than 0 or negative numbers, but all 10 of the soldiers apparently followed the suggestion implicit in this question and restricted their estimates to the range of 1 to 10. This certainly reduced the between-subjects variance of the magnitude estimates, and beyond this, it is hard to guess what the effects of this restriction might have been, compared with the usual use of a wider range of numbers.

2. Results

The arithmetic means of the magnitude estimates for the four machine guns for each of the eight questions asked are presented in table 8.

Eight repeated measures analyses of variance were carried out on the data from which the means in table 8 were computed. These analyses of variance are presented in table 9. From the data in table 9, it can be seen that the magnitude estimates differed significantly among the four machine guns for questions 2, 5, and 7. Referring to table 8, it can be seen that the MG-1A3 and M60 machine guns were considered easier to open than the RPK and PKM machine guns (question 2). It was considered easier to operate the charging handle on the MG-1A3 than on the other three machine guns (question 5). And using the safety was considered easier on the MG-1A3 and M60 machine guns than on the RPK and PKM machine guns (question 7).

3. Discussion

Use of the magnitude estimation technique appears to have been successful in measuring the soldiers' subjective responses to the four machine guns, in that statistically significant results were obtained for three of the eight questions. The results for these three questions were in accordance with the soldiers' informal opinions of the machine guns. As noted earlier, it is impossible to estimate the effects that the soldiers' use of a 1 to 10 scale may have had on these results. An objective, external criterion was available for only one of the eight questions--that concerning accuracy. Average horizontal and vertical miss-distances for the four machine guns at both 300 and 600 meters were obtained from the draft report of the MAD test, and are presented in table 10. Examination of these miss-distances shows

Table 8. Arithmetic Means of Magnitude Estimates for Four Machine Guns Entered on Data Collection Form

	MG-1A3	RPK	PKM	M60
1. The weapon was-- Accurate: big number Inaccurate: small number	6.2	6.2	6.6	8.0
2. The weapon was-- Easy to open: big number Hard to open: small number	8.3	7.3	7.1	8.5
3. Aiming from the bipod was-- Easy: big number Hard: small number	7.0	5.7	6.6	7.4
4. Firing from the bipod was-- Easy: big number Hard: small number	7.2	7.0	7.1	7.1
5. Operating the charging handle was-- Easy: big number Hard: small number	8.6	6.7	6.1	6.6
6. Squeezing the trigger was-- Easy: big number Hard: small number	8.1	7.3	7.0	8.1
7. Using the safety was-- Easy: big number Hard: small number	8.3	7.0	6.9	8.3
8. Overall, do you consider this weapon-- Good: big number Poor: small number	7.7	6.7	7.0	8.2

Table 9. Analysis of Variance of Magnitude Estimates for Four Machine Guns

1. The weapon was accurate--inaccurate?

Source Variance	SS	df	MS	F	p
Between subjects	128.43	9	--	--	--
Within subjects	116.12	30	--	--	--
Machine guns	22.33	3	7.44	2.13	n.s.*
Error	93.79	27	3.47	--	--
TOTAL	244.55	39			

2. The weapon was easy to open--hard to open?

Source Variance	SS	df	MS	F	p
Between subjects	179.70	9	--	--	--
Within subjects	38.60	30	--	--	--
Machine guns	13.95	3	4.65	5.11	<.01
Error	24.65	27	.91	--	--
TOTAL	218.30	39			

3. Aiming from the bipod was easy--hard?

Source Variance	SS	df	MS	F	p
Between subjects	229.76	9	--	--	--
Within subjects	68.81	30	--	--	--
Machine guns	16.52	3	5.51	2.84	n.s.
Error	52.29	27	1.94	--	--
TOTAL	298.57	39			

4. Firing from the bipod was easy--hard?

Source Variance	SS	df	MS	F	p
Between subjects	[illegible]	9	--	--	--
Within subjects	[illegible]	30	--	--	--
Machine guns	.22	3	.07	<1	n.s.
Error	53.61	27	1.99	--	--
TOTAL	232.17	39			

5. Operating the charging handle was easy--hard?

Source Variance	SS	df	MS	F	p
Between subjects	128.32	9	--	--	--
Within subjects	96.86	30	--	--	--
Machine guns	35.68	3	11.89	5.24	<.01
Error	61.18	27	2.27	--	--
TOTAL	225.18	39			

6. Squeezing the trigger was easy--hard?

Source Variance	SS	df	MS	F	p
Between subjects	138.50	9	--	--	--
Within subjects	55.11	30	--	--	--
Machine guns	9.28	3	3.09	[illegible]	n.s.
Error	45.83	27	1.70	--	--
TOTAL	193.61	39			

7. Using the safety was easy--hard?

Source Variance	SS	df	MS	F	p
Between subjects	118.48	9	--	--	--
Within subjects	57.80	30	--	--	--
Machine guns	18.87	3	6.29	4.37	<.025
Error	38.93	27	1.44	--	--
TOTAL	176.28	39			

8. Overall, do you consider this weapon good--poor?

Source Variance	SS	df	MS	F	p
Between subjects	143.99	9	--	--	--
Within subjects	67.50	30	--	--	
Machine guns	14.23	3	4.47	2.41	n.s.
Error	53.27	27	1.97	--	--
TOTAL	211.49	39			

LEGEND: SS Sum of Squares
df Degrees of Freedom
MS Mean Square
F F-Ratio
p Probability
n.s. Not significant, p>.05

no consistent pattern; none of the machine guns appear to be more accurate than any other machine gun. This lack of a consistent pattern in the objective accuracy data is consonant with the fact that there were no statistically significant differences in subjective judgments as to the accuracy of the four machine guns.

C. MAGNITUDE ESTIMATION OF DIFFICULTY OF MAN-PACK PORTABILITY COURSE

The Man-Pack Portability Course (MPPC) is the same 4-kilometer jungle course as was used in the PASGT test of helmets and vests mentioned earlier. Detailed descriptions of the development and intended

Table 10. Average Miss-Distances for Four Machine Guns

Machine Gun	Miss-Distances* 300 meters Horizontal	Vertical	600 meters Horizontal	Vertical
MG-1A3	-23.9	9.0	-54.8	13.5
RPK	-15.3	-17.7	13.9	-1.3
PKM	-14.8	28.7	-25.2	39.5
M60	-19.8	-10.6	-18.4	7.2

* Each average miss-distance is based on 500 rounds (50 rounds by each of 10 soldiers). Miss-distances are in inches. Negative values are to the left of, or below the bullseye; positive values are to the right of, or above the bullseye.

uses of this course may be found in Test Operations Procedure (TOP) 1-3-550 (1973), Williamson and Kindick (1974)[10] and Williamson and Kindick (1975)[11]. Groups of soldiers carried four different loads in order to introduce variation in difficulty of traversing the course. The soldiers were then asked to give magnitude estimates of the difficulty of several parts of the course.

1. Details of Experimental Procedures

Four groups of five soldiers each traversed the MPPC on each of 4 consecutive days, 5-8 September 1978. The four groups started the course at approximately 30-minute intervals, so they would not encounter each other on the course. Each group took 2 to 3 hours to traverse the course, and on each day the groups traversed the course generally between 0800 and 1200 hours. Each of the four groups of soldiers came from a different company of the 4th Battalion (Mech), 20th Infantry, stationed at Fort Clayton, Canal Zone (Companies A, B and C, and the Combat Support (CS) Company). Four persons from USATTC traversed the course with the soldiers, one with each group, serving as timers of the performances of the men on various parts of the course. Each of the soldiers carried his weapon (M-16) and a cartridge belt with two canteens, in addition to a pack which was loaded to one of four weights: 20-25 pounds, 30-35 pounds, 40-45 pounds, and 50-55

10 Williamson, R. L., and Kindick, C. M. Human Performance in the Tropics I: Man-Packing a Standard Load Over a Typical Jungle Course in the Wet and Dry Season.

11 Williamson, R. L., and Kindick, C. M. Human Performance in the Tropics II: A Pilot Study on Load-Carrying Test Methodology.

pounds. The loads carried, the order of traversing the course, and the assignment of timers to the groups were arranged in a balanced fashion, as shown in table 11. It can be seen in table 11 that each group carried each of the four different loads, traversed the course once in each of the four turns (1st, 2nd, 3rd, and 4th), and was assigned each of the four different timers.

Table 11. Experimental Design for Loads Carried, Order of Traversing the MPPC and Assignment of Timers

Order of Traversing MPPC	Days			
	5 Sep	6 Sep	7 Sep	8 Sep
1st	Co. CS 20-25 lbs (9.1-11.3 kg) Timer 4	Co. C 50-55 lbs (22.7-24.9 kg) Timer 4	Co. B 50-55 lbs (22.7-24.9 kg) Timer 4	Co. A 20-25 lbs (9.1-11.3 kg) Timer 4
2nd	Co. C 30-35 lbs (13.6-15.9 kg) Timer 3	Co. CS 40-45 lbs (18.1-20.4 kg) Timer 2	Co. A 40-45 lbs (18.1-20.4 kg) Timer 2	Co. B 30-35 lbs (13.6-15.9 kg) Timer 3
3rd	Co. B 40-45 lbs (18.1-20.4 kg) Timer 2	Co. A 30-35 lbs (13.6-15.9 kg) Timer 3	Co. CS 30-35 lbs (13.6-15.9 kg) Timer 3	Co. C 40-45 lbs (18.1-20.4 kg) Timer 2
4th	Co. A 50-55 lbs (22.7-24.9 kg) Timer 1	Co. B 20-25 lbs (9.1-11.3 kg) Timer 1	Co. C 20-25 lbs (9.1-11.3 kg) Timer 1	Co. CS 50-55 lbs (22.7-24.9 kg) Timer 1

The same five persons remained in each of the four groups for the 4 days, with the exception of some necessary substitutions. Of the total of 80 individual traverses of the MPPC (20 persons x 4 days), 10 were accomplished by substitutes. Thus the substitution rate was 12.5 percent. Twelve of the 20 soldiers who traversed the course on the first day were present and traversed the course on each of the 3 remaining days.

A traversal of the MPPC consists of several different parts, as follows:

--Forced March, 5,200 feet, group timed
--15-minute break
--Walk to Frijoles River, 3-minute break while crossing river
--Walk to beginning of Uphill Run (timer goes ahead to top of hill)
--Uphill Run, 300 feet, individually timed
--15-minute break

--Walk down hill to Frijoles River, 3-minute break
--Walk to marker 283
--5-minute break
--Walk to beginning of Double Time (timer goes ahead to end of Double Time Course)
--Double Time, 200 feet, individually timed
--5-minute break
--Walk to end of course

The Total Time required by each group to traverse the MPPC was recorded. To obtain a time for each group for the Normal Walk portions of the course, the group time for the Forced March, the sum of the individual times for the Uphill Run, the sum of the individual times for the Double Time, and 46 minutes of breaks were subtracted from the Total Time required by the group to traverse the course. The men in each group were identified by designations (A-1, A-2, . . ., A-5; B-1, . . ., B-5; C-1, . . ., C-5; CS-1, . . ., CS-5) written on strips of white engineer tape tied around their upper arms.

Both before they traversed the MPPC and after they had traversed the MPPC, the soldiers were asked to strip to their shorts to be weighed. Likewise, their canteens (without cup or cover) were weighed when full before traversing the course. Body weights were recorded to the nearest 10th of a pound, and canteen weights to the nearest ounce. Body weight after traversing the course was subtracted from body weight before traversing the course to obtain body weight loss. Likewise, weight of empty or partially empty canteens was subtracted from weight of full canteens to obtain the weight of water drunk by each soldier while traversing the course. Weight of water drunk was then added to body weight loss to obtain a measure of sweat loss for each soldier during traversal of the course. Finally, sweat loss was expressed as a percentage of initial body weight.

Thus, on each of the 4 days the following objective data were collected:

--Forced March time, group measure (N = 4), to nearest minute
--Uphill Run time, individual measure (N = 20), to nearest second
--Double-Time time, individual measure (N = 20), to nearest second
--Total Time, group measure (N = 4), to nearest minute
--Normal Walk time, group measure (N = 4), to nearest minute
--Percent Initial Body Weight Lost, individual measure (N = 20), to nearest 0.1 percent.

It is recognized that Total Time is not independent of the other times, because Total Time is a composite of the other scores.

After each group finished traversing the MPPC on each day, subjective data were gathered by asking each soldier to make magnitude estimates of the difficulty of the following parts of the course:

--Forced March
--River Crossing
--Uphill Run
--Double Time
--Walking Parts

Weather conditions were typical rainy season conditions for the Canal Zone, with frequent afternoon and evening rains. Thus the course, which is almost entirely under jungle canopy, was quite slippery and muddy on all 4 days. However, it did not rain while the groups were actually traversing the course, with the exception of the last 15 minutes of the traversal by the last group on the 4th day.

On the 1st day when the soldiers arrived at the site of the MPPC, they were given the following explanation of what they were to do:

> We're trying out a new method of measuring what you think about some job or task. We're going to ask you to go over what we call a Man-Pack Portability Course in the jungle. It's about 4 kilometers long and has several parts in it. The course is marked with yellow arrows nailed to trees. You may find some fallen trees across the course, but just go ahead and follow the arrows, climbing over or going around any obstacles. After you've gone over the course, I will ask you to tell me how difficult the parts of the course were by simply giving me a number for each part of the course; the more difficult the part of the course, the bigger the number you should give me.
>
> You will go over the Man-Pack Portability Course in groups of five men, one group from each Company. On some parts of the course we will time you as a group, and on other parts we will time each one of you separately. We're interested in how much weight you lose in sweat as you go over the course, so we'll weigh you before you go onto the course, and after you come off it. We will ask you to strip to your shorts when we weigh you, and we also want to weigh your canteens.
>
> Now, each company came out here with a set of five loads to carry on the course. Some of the loads are heavy and some are lighter. We'll switch these loads around each day, so that each of you will carry all four different loads, from light to heavy, over the 4 days we'll be out here. We will start today with the men from each company carrying the loads they brought with them.
>
> Here's a chart that tells you a little about what's in the Man-Pack Portability Course: first there's a 5,200-foot Forced March, nearly a mile. We time you as a group on this. Then you take a 15-minute break. Then you walk on down to the Frijoles river and take a 3-minute break here.

Please don't drink the river water because it's probably contaminated. After you cross the river, you walk part way up a hill, to the starting point of the 300-foot Uphill Run. Then, one at a time, you run as fast as you can, 300 feet up the hill. You'll be timed individually on this. At the top of the hill you take another 15-minute break. After the break you walk down the hill to the river where you take a 3-minute break. Again, don't drink the river water. Then you'll walk on along the course to Arrow No. 283, and take a 5-minute break here. You'll continue walking to the starting point of the 200-foot Double Time. Here, one at a time, you run as fast as you can for 200 feet over level ground in the jungle. We time you individually on this. After all of you in a group have completed the Double Time, you take a 5-minute break. Then you walk on back here to the end of the course.

Now let's get you lined up into four five-man groups by companies, and get you identified with some tape markers. We'll weigh you first, and then when you have put your uniforms back on, we'll put some tape markers on you. The group from Combat Support Company, carrying 20-25-pound packs, will be first. When you get back from the course, I will be asking you questions about how difficult you thought it was.

As the men in the first group were being weighed and fitted with tape markers on their arms, the following instructions were given to the four timers:

Your main job is to time the soldiers as they go through the Man-Pack Portability Course--time the performance events and the breaks, etc. Otherwise, you should leave matters such as setting the pace on the Forced March and walking parts of the course to the NCOIC of the group. Of course, you should be alert and see that none of the men get off the course.

Each of you should have a wristwatch, a stopwatch and a whistle. Here are the cards on which the times are recorded. Let me quickly go over them with you.

Group--should be A, B, C, or CS--the Company the men are from.

Date--be sure this is correct, otherwise the data will be confused.

Test Item--leave blank.

Start Time--wristwatch time when you start course, to the minute.

You'll be using the Yellow Course, so I have scratched out the marker numbers for the Red Course.

The Forced March goes from the start of the Course to Marker 145. You can use either your wristwatch or the stopwatch to time this, to the minute.

At Marker 145, the end of the Forced March, take a 15-minute break. Then walk at a normal easy pace down to the Frijoles River. Take a 3-minute break here, and warn the men not to drink river water or fill their canteens.

Then walk to the beginning of the Uphill Run, where you'll see the engineer tape on the trees. Here, have the men line up--1, 2, 3, 4 and 5, in order--and tell the NCOIC to have the men start the Uphill Run one at a time when you signal from the top of the hill. Then go up the hill to the end of the Uphill Run. Signal that you're ready with three blasts on the whistle, then one blast to start man no. 1--and start the stopwatch! When man no. 1 finishes the Uphill Run, record his time and then give one blast on the whistle as a signal for man no. 2 to start--and so forth. Now, let's check stopwatches to be sure they're wound and that you know how to operate the one you have.

After the last man in the group finishes the Uphill Run, take a 15-minute break, then walk at a normal pace down to the river. Take a 3-minute break at the river--again no drinking from the river or filling canteens--and then a normal walking pace to marker 283. Take a 5-minute break here, and then walk to the beginning of the Double Time, where you'll again see the engineer tape tied to the trees. Have the men line up--you go to the end of the Double Time course, and start and time them individually, as you did on the Uphill Run.

When the last man finishes the Double Time, take a 5-minute break, then walk at a normal pace to the end of the course. Record end time from your wristwatch, to the nearest minute, when the last man in the group reaches the end of the course.

As the group returned from the course on the 1st day, after they had been weighed and had their canteens weighed, the experimenter took them, one at a time, and gave them the following explanation of magnitude estimation:

I'd like you to tell me how difficult the various parts of the course were by simply giving me a number. If you thought a part of the course was very hard you should give me a big number. If you thought it was easy, you should give me a small number. Please do not give me any zeros or negative numbers, though. Now, think about the Forced March. If you

thought it was very hard, give me a big number. If you thought it was very easy, give me a small number.

What about crossing the river?
Uphill Run?
Double Time?
Walking parts of the course?

On succeeding days; 6, 7 and 8 September, after the soldiers had finished the course and been weighed, they were greeted with:

OK, you've been over the course and done this before. What about the Forced March today? Et cetera.

2. Analyses and Results: Magnitude Estimation

The loads carried by the soldiers as they traversed the MPPC were intended to influence their perception of the difficulty of the various parts of the MPPC, in the same manner as the actual lengths of the stimulus lines presented to subjects in the laboratory experiments were shown to determine their perceived line lengths. Magnitude estimation was used to measure the subjective variables in both cases; perceived line lengths in the laboratory experiments, and perceived difficulties of various parts of the MPPC in the field experiment now under consideration. Thus, it is reasonable to see how well the expected relationships between loads carried over the MPPC and the perceived difficulties of various parts of the MPPC can be represented by power functions. Of course, it cannot be expected that the relationships between loads carried and perceived difficulties of various parts of the MPPC will be represented so well by power functions (or any other exact functions, for that matter) as was found to be the case with lengths of stimulus lines and perceived line lengths in the laboratory experiments. As was pointed out at the beginning of this section, this is true because perceived difficulties of the various parts of the MPPC will clearly depend on a number of other physical, physiological and psychological variables, in addition to the independent variable manipulated in this field experiment--the loads carried by the soldiers as they traversed the MPPC.

To see how well power functions represent the relationships between loads carried over the MPPC and the perceived difficulties of various parts of the MPPC, as measured by magnitude estimation, both the loads carried over the MPPC and the geometric means of the magnitude estimates were converted to decibels and plotted on linear coordinates, as was done in laboratory experiments. The data used in these analyses are those provided by the 12 soldiers who traversed the MPPC and provided magnitude estimates of the difficulties of parts of the MPPC on all 4 days. Table 12 shows ratios and conversions to decibels of the stimulus loads carried, and of the five kinds of subjective responses--the perceived difficulties of the five parts of the MPPC.

Table 12. Ratios and Conversions to Decibels: Magnitude Estimates of Difficulty of MPPC

Stimuli

Loads Carried lbs (kg)	Ratios	Decibels
22.5 (10.2)	1.00	0.00
32.5 (14.7)	1.44	1.60
42.5 (19.3)	1.89	2.76
52.5 (23.8)	2.33	3.68

Response 1. Forced March

Geom Means of Mag Est	Ratios	Decibels
3.99	1.00	0.00
8.47	2.12	3.27
7.44	1.86	2.71
13.88	3.48	5.41

Response 2. River Crossing

Geom Means of Mag Est	Ratios	Decibels
1.99	1.00	0.00
3.38	1.70	2.30
3.61	1.81	2.59
3.80	1.91	2.81

Response 3. Uphill Run

Geom Means of Mag Est	Ratios	Decibels
8.74	1.00	0.00
13.46	1.54	1.88
19.08	2.18	3.39
21.32	2.44	3.87

Response 4. Double Time

Geom Means of Mag Est	Ratios	Decibels
3.95	1.00	0.00
6.75	1.71	2.33
5.22	1.32	1.21
12.54	3.18	5.02

Response 5. Walking Farts

Geom Means of Mag Est	Ratios	Decibels
3.74	1.00	0.00
4.96	1.33	1.23
6.48	1.73	2.39
9.09	2.43	3.86

When the decibel values of the geometric means of the magnitude estimates of the difficulties of each of the five parts of the MPPC are plotted in turn against the decibel values for the loads carried, the results obtained are those presented in figure 9. Slopes of best-fitting lines are generally in the vicinity of 1.00, and vary from .75 to 1.28. The fit of the lines to the points, and therefore of the corresponding power function to the data, ranges from quite good for the Uphill Run and Walking Parts of the MPPC, to rather poor for the Double Time. It should be noted that the ranges of both the stimulus variables--loads carried, and the response variables--magnitude estimates of the parts of the MPPC, are only 3 to 5 decibels, as contrasted with corresponding ranges of around 20 decibels in the laboratory experiments on magnitude estimation of line length. Thus, in this experiment, it was possible to investigate only relatively small portions of the relationships between loads carried and perceived difficulty of parts of the MPPC.

It may be of interest to know the results of an analysis of differences between arithmetic means of the magnitude estimates. Table 13 presents the arithmetic means of the magnitude estimates of difficulty, for the four different loads carried and for the five different parts of the MPPC. An analysis of variance was performed of the data on which the means of table 13 are based. This analysis was a two factor analysis with repeated measures on both factors. The results of this analysis are shown in table 14. Though the means in table 13 appear to differ widely, this analysis of variance shows that the differences are not statistically significant; either for the parts of the MPPC, the loads carried, or for the interaction between these two factors.

The between-subjects variability of these magnitude estimates was large, with one subject using only the range of 1 to 4, and 11 of 12 subjects using ranges between 1 and 60, while one subject used a range of 10 to 1,000. The extreme values of the magnitude estimates of this one subject obviously had a large effect on most of the arithmetic means in table 13. However, if the geometric means in table 12 are compared with the corresponding arithmetic means in table 13, it may be seen that the geometric means were much less affected by the extreme values of the magnitude estimates made by this one subject.

3. <u>Analyses and Results: Performance Data</u>

The times required to traverse the various parts of the MPPC and the percentage of initial body weight lost in perspiration are objective, performance data yielded by this field study. It will be of interest to analyze and compare the results with those from the magnitude estimation data. For the Uphill Run, Double Time and Percent Initial Body Weight Lost variables, individual data are available; for the Forced March, Total Time and Normal Walk variables, the data are for groups. In the individual performance data, the data are for 20

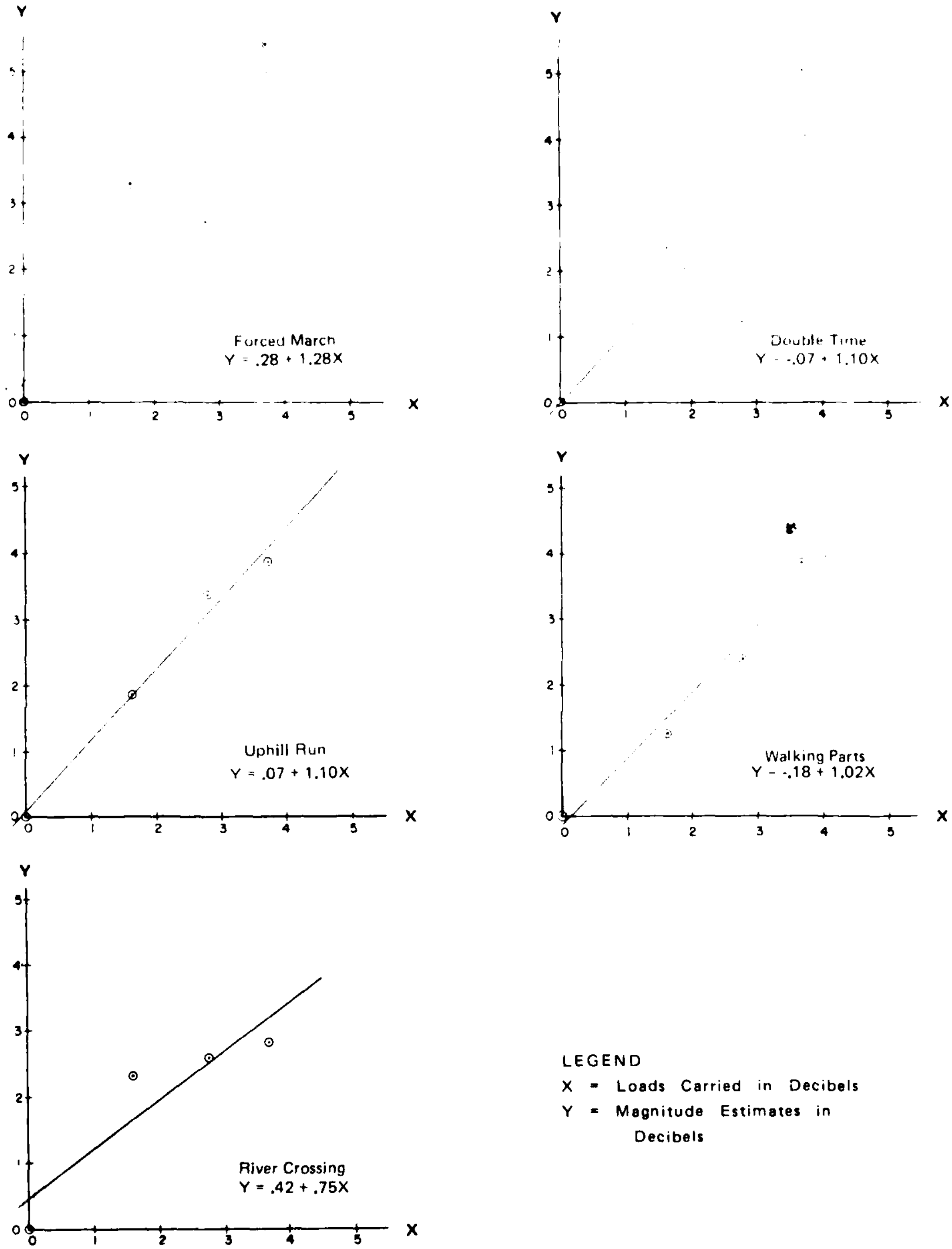

Figure 9. Magnitude Estimates of Difficulty of MPPC: Twelve Subjects, One Trial with Each of Four Loads Carried.

Table 13. Arithmetic Means of Magnitude Estimates of Difficulty*

Parts of MPPC	Loads Carried: 20-25 lbs (9.1-11.3 kg)	30-35 lbs (13.6-15.9 kg)	40-45 lbs (18.1-20.4 kg)	50-55 lbs ([illegible]-24.9 kg)
Forced March	8.00	39.75	[illegible].08	[illegible]
River Crossing	[illegible].50	11.08	[illegible].4	[illegible]
Uphill Run	50.17	95.1[illegible]	[illegible].9[illegible]	[illegible]04.5[illegible]
Double Time	5.50	46.4[illegible]	21.0[illegible]	[illegible].4[illegible]
Walking Parts	29.1[illegible]	6[illegible].4[illegible]	[illegible]	[illegible]

*Data on 12 soldiers who carried all four loads over the MPPC.

Table 14. Analysis of Variance of Magnitude Estimates of Difficulty*

Source of Variance	SS	df	MS	F	p
Between Subjects	3,529,975.65	11			
Within Subjects	2,227,199.85	228			
Parts of Course	159,960.48	4	39,990.12	1.44	[illegible]**
Error-P	1,220,646.62	44	27,741.97		
Loads	61,042.58	3	20,347.53	1.3[illegible]	[illegible]
Error-L	503,298.07	33	15,251.46		
P X L	29,936.19	12	[illegible],494.68	1.[illegible]	[illegible]
Error-P X L	252,215.91	132	1,911.48		
TOTAL	5,757,175.50	239			

*Data on 12 soldiers who carried all four loads over the MPPC.

**Not significant, $p > .05$.

subjects on each of the 4 days; i.e., data from substitutes are included. This was done because, in the performance events, the data obtained from substitutes were much more similar to those obtained from the regular subjects, than was the case with the magnitude estimation data where a substitute often would use a very different range of numbers from that used by the regular subject whom he replaced.

The arithmetic means of the performance variables for the four loads carried over the MPPC are presented in table 15. An analysis

variance was carried out on the data for each performance variable. These analyses were of two kinds. For the individual data, two-factor analyses were carried out with repeated measures on one of the factors. The between-subjects factor is the groups factor; i.e., the four groups in which the subjects traversed the MPPC. The within-subjects factor is the loads-carried factor. For the group data, the analyses were single-factor analyses with repeated measures on the loads-carried factor.

In the analyses of the individual data, the groups factor is not of any particular interest; i.e., it is expected that significant differences between groups will be found. The groups factor is included in the analyses in order to make the test more sensitive for loads-carried effects.

Table 15. Arithmetic Means of Performance Variables

Performance Variables	20-25 lbs (9.1-11.3 kg)	30-35 lbs (13.6-15.9 kg)	40-45 lbs (18.1-20.4 kg)	50-55 lbs (22.7-24.9 kg)	N Groups	N Persons
Forced March (Min)	[illegible]	[illegible]	34.75	32.50	4	
Uphill Run (Sec)	61.55	66.45	[illegible]	87.40		20
Double Time (Sec)	19.00	27.50	19.20	19.60		20
Total Time (Min)	[illegible]	131.50	159.50	152.75	4	
Normal Walk (Min)	[illegible]	[illegible]	70.75	65.25	4	
Initial Body Weight Lost %	2.43	2.49	2.57	2.48		20

The analyses of variance of the individual data (Uphill Run, Double Time and Percent Initial Body Weight Lost) are presented in table 16. It can be seen that differences between average times required to cover the Uphill Run and the Double Time events, for different loads carried were significant, but differences between averages of Percent Initial Body Weight Lost were not. These analyses also show significant interactions between groups and loads carried on both the Uphill Run and Double Time events. This means that the relationship between loads carried and performance on these events is significantly different for at least one group, as compared with the other groups. To

Table 16. Analyses of Variance of Individual Scores on MPPC

Uphill Run

Source of Variation	SS	df	MS	F	p
Between Subjects	14,624.14	19			
A (Groups)	7,744.04	3	2581.35	6.00	<.01
Subjects within Groups	6,880.10	16	430.01		
Within Subjects	44,938.25	60			
B (Loads)	8,244.74	3	2748.25	8.26	<.001
AB	20,715.21	9	2301.69	6.91	<.001
B subjects within Groups	15,978.30	48	332.88		
TOT	59,562.39	79			

Double Time

Source of Variation	SS	df	MS	F	p
Between Subjects	1,315.55	19			
A (Groups)	1,068.15	3	356.05	23.03	<.001
Subjects within Groups	247.40	16	15.46		
Within Subjects	4,634.00	60			
B (Loads)	1,020.55	3	340.18	46.41	<.001
AB	3,261.65	9	362.41	49.44	<.001
B X Subjects within Groups	351.80	48	7.33		
TOTAL	5,949.55	79			

Percent Initial Body Weight Lost

Source of Variation	SS	df	MS	F	p
Between Subjects	25.35	19			
A (Groups)	6.40	3	2.13	1.81	n.s*
Subjects within Groups	18.95	16	1.18		
Within Subjects	29.43	60			
B (Loads)	0.19	3	0.06	<1.00	n.s
AB	9.21	9	1.02	2.43	n.s
B X Subjects within Groups	20.03	48	0.42		
TOTAL	54.78	79			

*Not significant, p >.05

investigate this, arithmetic means for each of the four groups were computed, for each load carried. For the Uphill Run, these means are presented in table 17. Figure 10 shows a plot of these means and from this plot it is immediately obvious that the performance of the CS Company group when carrying 50- to 55-pound loads (on the last of the 4 days) is very different than would be expected from the performances of the other three groups (and of the CS group when carrying other loads). When the data were analyzed and this became apparent, an inquiry was started to try to determine reasons for this deviant performance. No clear-cut explanation was found, but the following reasons were suggested, somewhat speculatively.

(a) The men of the CS Company, who worked in a variety of maintenance and support jobs, were not as well conditioned physically as the men in the other companies, who had trained extensively in the jungle.

(b) The men in the other companies were challenged by the MPPC, and heavier loads appeared only to arouse greater determination in them; while the men of the CS Company regarded traversing the MPPC as almost a punishment, and the 50- to 55-pound loads were considered as the "last straw." In other parts of the MPPC, when carrying the 50- to 55-pound loads, the CS group had high (but not the highest) times, but the Uphill Run was much more demanding than the other parts of the MPPC, and it is possible that the morale of this group suffered considerably at this point.

The arithmetic means on the *Double Time* event of the four groups, for each load carried, are presented in table 18. Figure 11 shows a plot of these means and again the deviant performance is perfectly obvious. This time it was the Company C group when carrying 30 to 35 pounds (on the first of the 4 days) that performed very differently than the other groups, and also very differently than they (Company C) did when carrying other loads. Inquiry revealed a clear-cut reason for this different performance; the timer who accompanied them on this first day remembered that this group had worn combat boots instead of jungle boots on this one day, and had slipped a good deal while traversing the course. Detailed examination of the data for other parts of the MPPC showed that on this day the Company C group had higher times than any of the other groups did while carrying 30 to 35 pounds, and that the Company C group had a higher average score for Percent Initial Body Weight Lost on this day than they did on any of the other 3 days. The Company C group average for Percent Initial Body Weight Lost was also considerably higher on this day than the averages for the other three groups were on any of the 4 days. Thus it is clear that the relatively poor traction of the combat boot had a considerable effect on the performance of this group, who had inadvertently worn them on this one day. Why it should have affected their performance on the Double Time event so much more than it did on other parts of the MPPC is not clear, unless it is because the Double Time event is near the end of the MPPC, and the group may have been very tired at that point.

Table 17. Uphill Run: Arithmetic Means for Four Groups

Groups	Loads Carried 20-25 lbs	30-35 lbs	40-45 lbs	50-55 lbs
A	62.80 sec	62.80 sec	61.40 sec	60.60 sec
B	52.40 sec	62.40 sec	68.00 sec	76.00 sec
C	55.20 sec	77.40 sec	70.60 sec	66.40 sec
CS	75.80 sec	63.20 sec	59.80 sec	148.00 sec

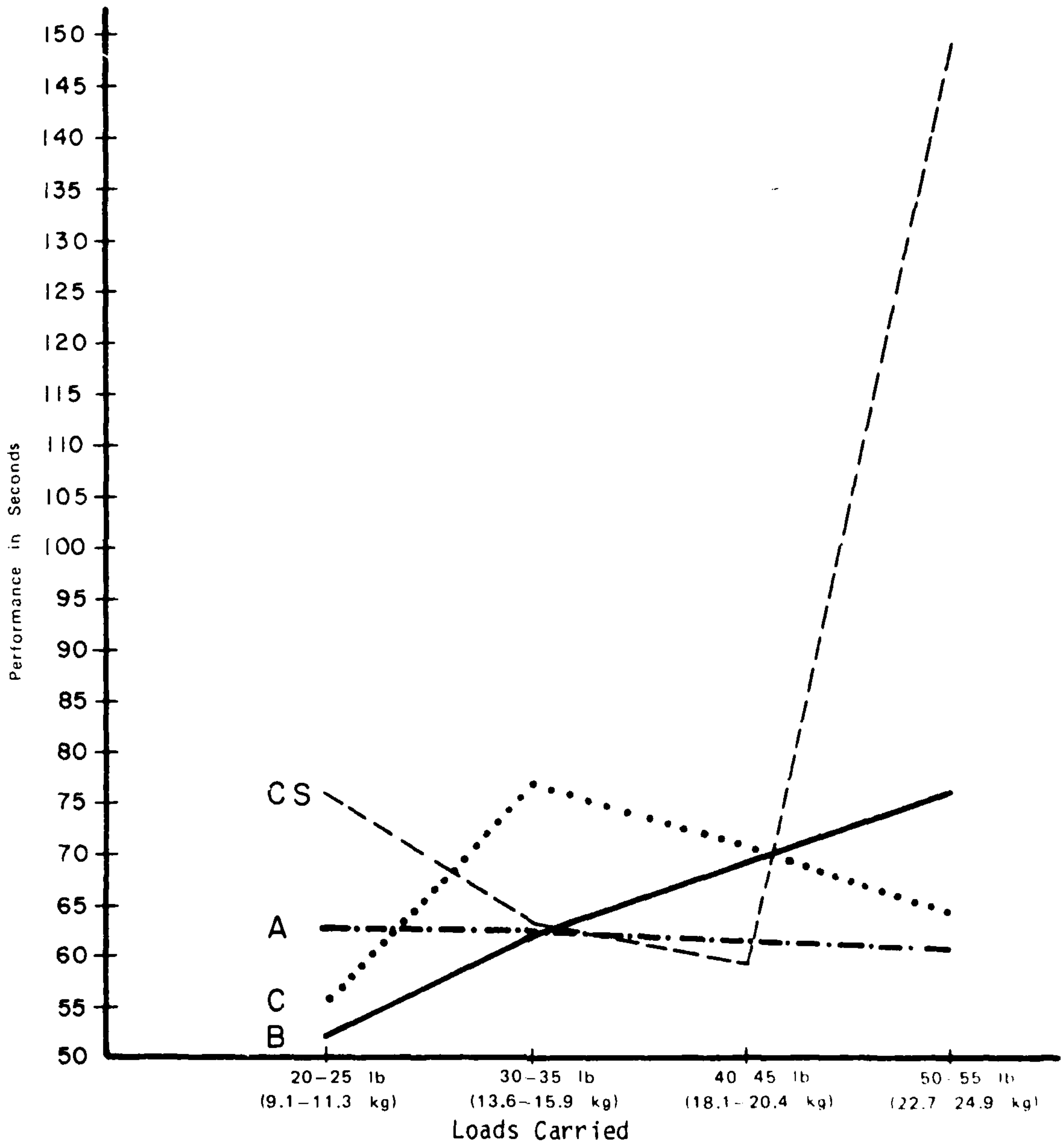

Figure 10. Uphill Run: Loads Carried, Arithmetic Means for Four Groups Plotted.

Table 18. Double Time: Arithmetic Means for Four Groups

Groups	Loads Carried 20-25 lbs	30-35 lbs	40-45 lbs	50-55 lbs
A	17.00 sec	17.20 sec	17.20 sec	18.80 sec
B	19.80 sec	19.60 sec	18.60 sec	18.20 sec
C	15.20 sec	52.00 sec	21.00 sec	20.40 sec
CS	24.00 sec	21.20 sec	20.00 sec	21.00 sec

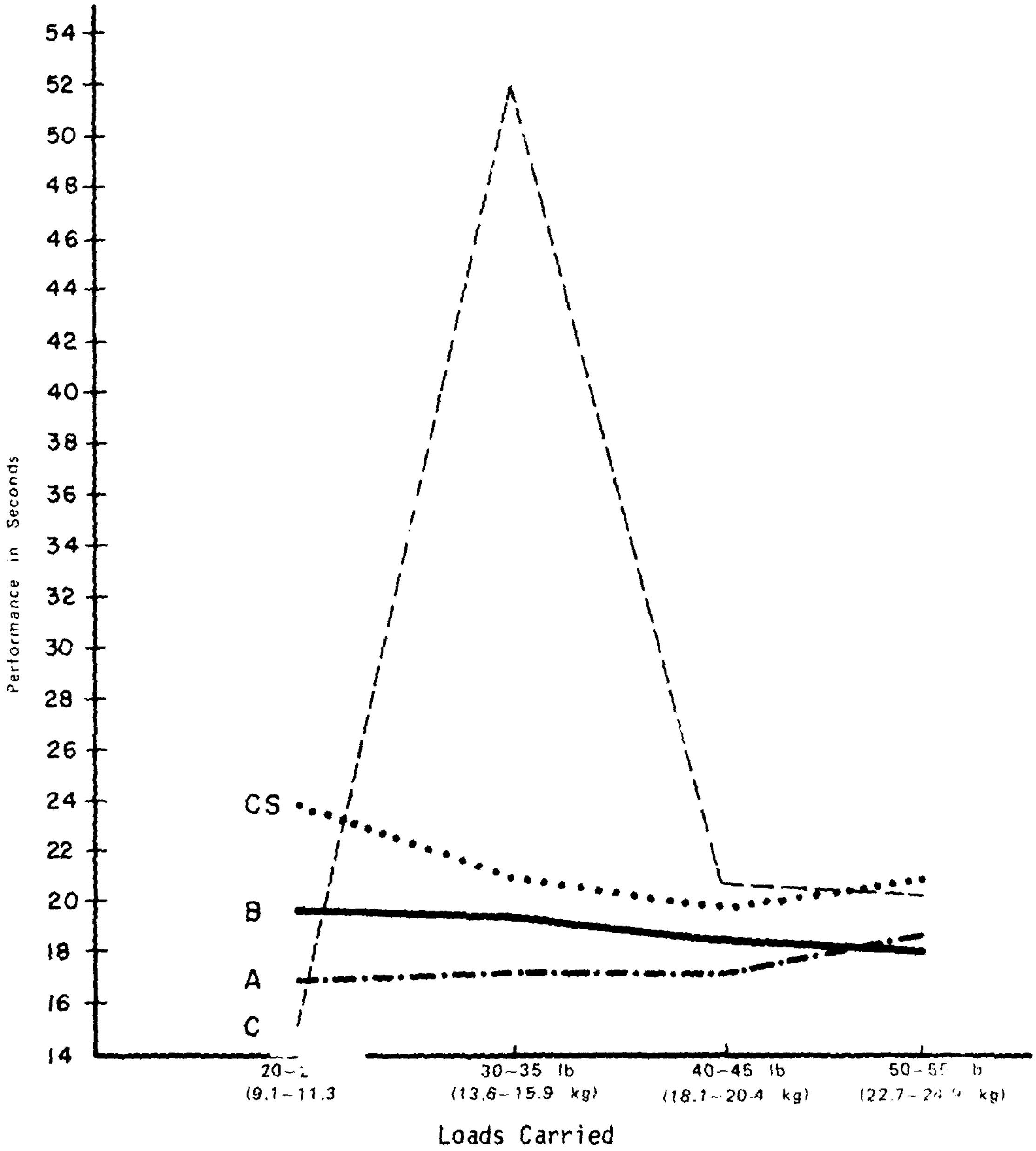

Figure 11. Double Time: Loads Carried, Arithmetic Means for Four Groups Plotted.

The analyses of variance for the group data (Forced March, Total Time and Normal Walk variables) are presented in table 19. These analyses show that the differences between average times for the groups on the Forced March, Total Time and Normal Walk variables, when carrying different loads, all differed significantly.

Table 19. Analyses of Variance for Group Scores on MPPC

Forced March

Source of Variance	SS	df	MS	F	p
Between Groups	32.75	3	--	--	--
Within Groups	357.00	12	--	--	--
Loads	207.25	3	69.08	4.15	<.05
Error	149.75	9	16.64		
TOTAL	389.75	15			

Total Time

Source of Variance	SS	df	MS	F	p
Between Groups	95.19	3	--	--	--
Within Groups	4821.75	12	--	--	--
Loads	2898.19	3	966.06	4.52	<.05
Error	1923.56	9	213.73		
TOTAL	4916.94	15			

Normal Walk

Source of Variance	SS	df	MS	F	p
Between Groups	201.50	3	--	--	--
Within Groups	2286.50	12	--	--	--
Loads	1381.00	3	460.33	4.58	<.05
Error	905.50	9	100.61		
TOTAL	2488.00	15			

4. Discussion

In summary, magnitude estimates of the difficulty of the Uphill Run and Walking Parts of the MPPC are fitted quite well by power functions of loads carried, with exponents of approximately 1.00. The fit for magnitude estimates of the difficulty of River Crossing is less satisfactory, and those for Forced March and Double Time are definitely poorer. Keeping in mind the expectations (discussed at the beginning of this section) that perceived difficulty of traversing parts of the MPPC would be affected by several variables in addition to the loads-carried variable, these results are a reasonably satisfactory outcome of this part of the field experiment. It is entirely possible that magnitude estimation did measure quite well the soldiers' subjective perceptions of the difficulty of various parts of the MPPC, and

that these perceptions were in some cases affected by other factors so that the relationships between loads carried and the magnitude estimates deviated from the expected power functions.

The failure of magnitude estimates of difficulty of the parts of the MPPC to show significant relationships with loads carried, when tested by analysis of variance, is probably because the four loads used in this study cover a limited segment of the possible range of loads. A study using both lighter and heavier loads than those used in this study would almost certainly show a significant relationship between loads carried and magnitude estimates of difficulty.

Analyses of the performance data did show that the loads-carried factor was significant for all variables except Percent Initial Body Weight lost. However, on the Uphill Run and Double Time variables, where individual data permitted analysis of the Groups by Loads interaction, it appeared that uncontrolled factors--such as the wearing of combat boots instead of jungle boots by the Company C group on one day and perhaps differences in physical conditioning and attitudes between the soldiers from the line companies (A, B and C) and those from the CS Company--accounted for substantial variations in performance. Examination of table 15 shows that for none of the performance variables was there a reasonably regular increase in means from lighter to heavier loads. Thus, the statistical significance of the loads carried factor in these analyses is probably due to uncontrolled factors, such as those described above, rather than reliable effects of the various loads carried.

SECTION VI. DISCUSSION AND CONCLUSIONS

Magnitude estimation was selected as the most promising response mode for evaluation in human factors settings, after a survey of cross-modality matching studies, because of its convenience and because of universal familiarity with the number scale. In a series of laboratory studies of magnitude estimation of line length, it was found that the data obtained by magnitude estimation were very well fitted by a power function, and that the exponent of the power function which best fitted the data was approximately 0.91 to 0.94. These results compared very well with those reported in the literature (Stevens, 1975); thus the experimenter was assured that the technique was being applied correctly in the laboratory studies. The data on magnitude estimation of line length were examined for evidence of the stability (reliability) of measurement and it was found that stability of measurement (at the level of group means with N = 12) was quite satisfactory. Particularly impressive were the intraclass correlation coefficients of 0.94 for the confirmatory experiment on magnitude estimation of line length.

Magnitude estimates for a group of subjects were found to vary widely with most of the estimates in a lower range of perhaps 1 to 10 up to 1 to 60, and a few estimates using much larger numbers. This is what would be expected with measurement on a ratio scale, instead of on an ordinal scale as yielded by other methods of measuring subjective variables. For example, consider the following geometric series: 1, 2, 4, 8, 16, 32, 64, 128, 256, in which the ratio of each term to the preceding one is constant and equals 2 in this case. When sampling from a population of values with a distribution of this kind, it would be expected that most values sampled would be in the lower range, with a few much higher values, as was found to be the case with magnitude estimates. That the geometric mean, rather than the arithmetic mean, is the appropriate measure of central tendency for data of this kind is shown by the fact that the geometric mean of such a series of values will be the middle value in the series--16 in the example above. The arithmetic mean will be unduly influenced by the few extremely high values, and is 56.78 in the example above.

It is not likely that direct evidence can be obtained that will make it intuitively obvious that subjective measurement by magnitude estimation is measurement on a ratio scale. Because the object of measurement is subjective, it is concealed within the private mental processes of each individual. In the future, sufficient information may be available on neural functioning so that the problem can be approached from this direction. But at present, it is not intuitively obvious that the units of the subjective scale are equal and that there is a natural zero point on this scale. Nevertheless, the evidence that magnitude estimates are distributed as a geometric series, cited above, points in the direction of the subjective scale being a ratio scale.

On the practical side, magnitude estimates were easily gathered in the preceding section. Of course, it is necessary that the persons who are to make the magnitude estimates understand clearly the variable on which they are to provide data: comfort, difficulty, confidence, or whatever it may be. And they should also understand that they may use any numbers they wish (except for zero and negative numbers); i.e., they should not restrict themselves to any preconceived range, such as 1 to 10.

Also, it should be clearly understood by the experimenter or tester that suitable experimental designs and methods of statistical analysis should be used. Generally, the principles of experimental design, such as suitable sampling, control and randomization procedures, apply when using magnitude estimation, just as they do in any other test or investigation. It is not definite at this point that the usual arithmetic mean and variance estimating statistical methods are entirely suitable for use with magnitude estimates. They were used in this study because no procedures were available for testing for differences between geometric means.

In the field studies described in the preceding section, magnitude estimation yielded results that generally made sense and agreed with available corroborating information. Differences in magnitude estimates between treatment groups were sometimes clear-cut, as in the comparisons of the helmets and vests, and sometimes less so, as in the comparisons of machine guns.

This study has clearly shown that magnitude estimation can be used to measure subjective variables in human factors evaluations. It appears to be superior to the rating and ranking methods used for many years, in that it achieves measurement on a ratio scale, as compared with the ordinal scale measurement achieved by the rating and ranking methods.

The author suggests that magnitude estimation be used in human factors evaluations, along with the usual rating and ranking methods of measuring subjective variables, so that the practical usefulness and validity, in terms of present day practices and standards, may be better assessed. When it is possible to compare the results of magnitude estimation with objective measures of performance that would be expected to be related to subjective variables, this should be done, since objective measures of performance provide a better basis for evaluation of magnitude estimation than do the usual rating and ranking methods of measuring subjective variables. Further empirical and theoretical investigations should be carried out on the sampling distributions of magnitude estimates and on the suitability (or unsuitability) of the usual arithmetic mean and variance estimation statistical methods for use with magnitude estimates. If necessary, statistical testing procedures should be developed for differences between geometric means.

APPENDIX A. BIBLIOGRAPHY

Klein, David, "Social Aspects of Exposure to Highway Crash," Human Factors, 18(3), Santa Monica, CA: Human Factors Society, Inc., 1976, pp. 211-219.

Williamson, R. L., and Dobbins, D. A., A New Approach Toward Quantifying Subjective Test Data, US Army Tropic Test Center Technical Report No. 7705001, TECOM Project No. 7-CO-IL7-TT1-001, AD No. A047838, Fort Clayton, CZ, May 1977.

TECOM Pamphlet 602-1, Vol 1, Man-Materiel Systems Questionnaire and Interview Design (Subjective Testing Techniques), 25 July 1975.

Stevens, S. S., ed., "Mathematics, Measurement and Psychophysics," Handbook of Experimental Psychology. New York: Wiley and Sons, 1951.

Stevens, S. S., Psychophysics. New York: Wiley and Sons, 1975.

Winer, B. J., Statistical Principles in Experimental Design. 2nd ed. New York: McGraw-Hill, 1971.

Shrout, P. E., and Fleiss, J. L., "Intraclass Correlations: Uses in Asssessing Rater Reliability, Psychological Bulletin, 86(2), Washington: American Psychological Association, 1979, 420-428.

Haverland, E. M.; Novak, C. A.; Johnson, R. L., Jr.; Williamson, R. L.; and Kindick, C. M., Development Test II of Personnel Armor System for Ground Troops (PASGT). US Army Tropic Test Center Report No. 7711001, TECOM Project No. 8-EI-525-000-27, AD No. B024817L, Fort Clayton, CZ, November 1977.

Test Operations Procedure (TOP) 1-3-550, Man-Pack Portability Testing in the Tropics, US Army Test and Evaluation Command, January 1973.

Williamson, R. L., and Kindick, C. M., Human Performance in the Tropics I: Man-Packing a Standard Load Over a Typical Jungle Course in the Wet and Dry Seasons. US Army Tropic Test Center Technical Report No. 7409002, TECOM Project No. 9-CO-049-000-004, AD No. A017274, Fort Clayton, CZ, September 1974.

Williamson, R. L., and Kindick, C. M. Human Performance in the Tropics II: A Pilot Study on Load-Carrying, Test Methodology. US Army Tropic Test Center Technical Report No. 7511001, TECOM Project No. 7-CO-IL6-TT1-001, AD No. A025460, Fort Clayton, CZ, November 1975.

APPENDIX B. DISTRIBUTION LIST

TECOM Project No. 7-CO-IL9-TT1-001
Magnitude Estimation: A New Method for Measuring Subjective Test Variables

Addressee	Final Report
Commander	
US Army Test and Evaluation Command	
ATTN: DRSTE-AD	1
DRSTE-AD-M	3
Aberdeen Proving Ground, MD 21005	
Commander	1
US Army Materiel Development and Readiness Command	
5001 Eisenhower Avenue	
Alexandria, VA 22333	
HQDA (DAMA-ARZ-B)	1
Washington, DC 20314	
HQDA (DAEN-ZA)	1
Forrestal Bldg	
Washington, DC 20314	
Assistant Secretary of Army (R&D)	
ATTN: Dr. Yore	1
Washington, DC 20310	
Associate for Documentation and Accession Psychological Abstracts	1
American Psychological Association	
1200 Seventeenth Street, NW	
Washington, DC 20036	
Institute for Applied Technology	1
National Bureau of Standards	
Washington, DC 20234	
Director	2
The Army Library	
The Pentagon, Rm 1A526	
Washington, DC 20310	
Director	
Naval Research Laboratory	
ATTN: Code 2627	1
Washington, DC 20375	
Smithsonian Institution	1
Office of the Assistant Secretary for Science (SI 114)	
1000 Jefferson Drive, SW	
Washington, DC 20560	
Commander	
US Army Aircraft Development Test Activity	
ATTN: Tech Director	1
Fort Rucker, AL 36362	
Commander	
US Army Electronic Proving Ground	
ATTN: STEEP-MT-I	1
Fort Huachuca, AZ 85613	

Addressee	Final Report
Commander	
US Army Yuma Proving Ground	
ATTN: STEYP-MMI	1
STEYP-MTD (Technical Library)	1
STEYP-MTS	1
Yuma, AZ 85364	
Commander	
Naval Training Equipment Center	
ATTN: DRCPM-TND	1
Orlando, FL 32813	
President	
US Army Infantry Board	
ATTN: Tech Director	1
Fort Benning, GA 31905	
Commander	
US Army Armament Materiel Readiness Command	
ATTN: DRSAR-RDM	1
Rock Island, IL 61201	
President	
US Army Armor and Engineer Board	
ATTN: Tech Director	1
Fort Knox, KY 40121	
Commander	
US Army Natick Research and Development Command	
ATTN: DRDNA-E	5
DRDNA-UBS	1
Natick, MA 01760	
Commander	
US Army Research Institute of Environmental Medicine	
ATTN: SGRD-VE-ADJ	1
Natick, MA 01760	
Director	
US Army Human Engineering Laboratory	
ATTN: DRXHE-D	1
DRXHE-PC	1
Aberdeen Proving Ground, MD 21005	
Director	
US Army Materiel Systems Analysis Activity	
ATTN: DRXSY-D	2
DRXSY-HEL	3
Aberdeen Proving Ground, MD 21005	

Addressee	Final Report
President US Army Airborne, Communications and Electronics Board ATTN: STEBG-TD Fort Bragg, NC 28307	1
Commander US Army Research Office ATTN: Information Processing Office P.O. Box 12211 Research Triangle Park, NC 27709	1
Commander US Army Armament Research and Development Command ATTN: DRDAR-DRG DRDAR-M110E2 Picatinny Arsenal Dover, NJ 07801	1 1
Chief US Army Human Engineering Laboratory ATTN: DRXHE-AR (Spine) ARRADCOM Bldg 333 Dover, NJ 07801	1
Commander US Air Force Test and Evaluation Center ATTN: AFTEC/XR Kirtland AFB, NM 87115	1
Commander US Army White Sands Missile Range ATTN: STEWS-TE White Sands, NM 88002	1
President US Army Field Artillery Board ATTN: ATZR-BDTD Fort Sill, OK 73503	1
President US Army Air Defense Board ATTN: ATZC-D-MT Fort Bliss, TX 79916	1
Commander TRADOC Combined Arms Test Activity ATTN: TCATA-OP Fort Hood, TX 76544	1
Commander US Army Dugway Proving Ground ATTN: STEDP-MT-L STEDP-SC Dugway, UT 84022	1 1
Director Defense Technical Information Center ATTN: DTIC-T Alexandria, VA 22314	2

Addressee	Final Report
Chief of Naval Research 800 N. Quincy Street Arlington, VA 22309	1
Director US Army Research Institute for the Behavioral and Social Sciences Reference Service 5001 Eisenhower Avenue Alexandria, VA 22333	1
Commander US Army Operational Test and Evaluation Agency ATTN: DACS-TEO-N 5600 Columbia Pike Falls Church, VA 22041	1
Commander US Army Tropic Test Center ATTN: STETC-MTD STETC-MTD (TIC) STETC-MTD (Tech Ed) APO Miami 34004	30 2 2

F

6